

(12) INTERNATIONAL APPLICATION PUBLISHED
UNDER THE PATENT COOPERATION TREATY (PCT)

(19) WORLD INTELLECTUAL PROPERTY ORGANIZATION
INTERNATIONAL BUREAU

(43) INTERNATIONAL PUBLICATION DATE	(10) INTERNATIONAL PUBLICATION NUMBER
13 September 2001(13.09.2001)	WO 01/67299 A1

(51) INTERNATIONAL CLASSIFICATION: G06F 17/30, C07K 7/026, 14/435, G01N 33/68

(21) INTERNATIONAL APPLICATION NUMBER: PCT/JP01/01846

(22) INTERNATIONAL APPLICATION DATE: 09 March 2001 (09.03.2001)

(25) LANGUAGE Japanese

(26) Japanese

(30) PRIORITY DATA: TOKUGAN 2000-72485 10 March 2000 (10.03.2000) JP

(71) APPLICANT:

DAIICHI PHARMACEUTICAL CO., LTD. [JP/JP]; 14-10, Nihonbashi 3-chome, Chuo-ku, Tokyo
103-8234 (JP).

FUJITSU LIMITED [JP/JP]; 1-1, Kamiodanaka 4-chome, Nakahara-ku, Kawasaki-shi, Kanagawa
211-8588 (JP).

(72) INVENTOR; AND INVENTOR/APPLICANT (US ONLY):

DOI, Hirofumi [JP/JP]; 29-4-515, Natsumi 5-chome, Funabashi-shi, Chiba 273-0865 (JP).

SUZUKI, Atsushi [JP/JP]; c/o DAIICHI PHARMACEUTICAL CO., LTD., Tokyo R & D Center,
16-13, Kitakasai 1-chome, Edogawa-ku, Tokyo 134-8630 (JP).

(74) AGENT:

SHOJI, Takashi, et al; 6F, SN Iwamotocho Bldg., 2-10, Iwamotocho 3-chome, Chiyoda-ku, Tokyo
101-0032 (JP).

(81) DESIGNATED STATES (NATIONAL): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ,
CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID,
IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZW

(84) DESIGNATED STATES (REGIONAL): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ,
TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)

Published

-- with international search report

METHOD FOR PREDICTING PROTEIN-PROTEIN INTERACTIONS

FIELD OF THE INVENTION

The present invention relates to a prediction for protein-protein interactions, a method and a device therefor, and proteins obtained using the method and device.

BACKGROUND OF THE INVENTION

Many proteins carry out their function by interacting with other proteins or the same protein. Thus, it is important to elucidate protein-protein interactions in the development of pharmaceuticals, the breeding in agriculture, and the like. Notably, along with the progress of genome analysis and cDNA analysis of various organisms including pathogenic microorganisms, the number of genes newly found and proteins encoded thereby whose functions are not known is rapidly increasing. Elucidating protein-protein interactions may permit one to predict the function of a protein whose function is not known.

A conventional method that has been used to screen for a protein interacting with a certain protein so as to elucidate their interactions, is the so-called two-hybrid system (Field, S. The two-hybrid system to detect protein-protein interaction. METHODS: A Companion to Meth. Enzymol., 5, 116-124, 1993). However, the two-hybrid system is a screening-based experiment, whose operation is complicated and time-consuming. Also, the number of proteins obtained is lower than that expected. In addition, this method has a disadvantage in that the results depend on the quality of the cDNA library used. In other words, this method has the risk that a gene encoding a protein interacting with a certain protein is not contained in the cDNA library used.

On the other hand, protein databases based on genome analysis and cDNA analysis have been enhanced, such that a method has also been adopted wherein a protein

complex in a cell is subjected directly to MALDI-TOF mass spectrometry, followed by searching in the database for a fragment of the amino acid sequence thereof (Yates, JR 3rd, J. Mass Spectrom. 33, 1-19, 1998; Humphrey-Smith, I., et al., Electrophoresis, 18, 1217-1242; Kaufmann, R., 1995, J. Biotechnol., 41, 155-175, 1997). This method gives information concerning proteins that form a complex, but does not give any information concerning the protein-protein interaction. Thus, it must be experimentally confirmed which proteins interact with each other.

SUMMARY OF THE INVENTION

In one embodiment, the present invention relates to a method for predicting a protein or polypeptide (B) that interacts with a specific protein or polypeptide (A), wherein the method is characterized by comprising:

- 1) decomposing the amino acid sequence of protein or polypeptide (A) into a series of oligopeptides having a pre-determined length as sequence information;
- 2) searching, within a database of protein or polypeptide amino acid sequences, for a protein or polypeptide (C) comprising an amino acid sequence for each member of the series or for a protein or polypeptide (D) comprising an amino acid sequence homologous to an amino acid sequence of each member of the series;
- 3) carrying out local amino acid sequence alignment between said protein or polypeptide (A) and the detected protein or polypeptide (C) or detected protein or polypeptide (D); and
- 4) predicting whether the detected protein or polypeptide (C) and/or protein or polypeptide (D) is a protein or polypeptide (B) that interacts with the protein or polypeptide (A) based on the results of the local amino acid sequence alignment and a value calculated from a frequency of an amino acid and/or a frequency of said oligopeptides in said amino acid sequence database.

One embodiment of the present invention may be the above-mentioned method for prediction wherein the oligopeptide is 4-15 amino acids in length.

In addition, in another embodiment, the present invention relates to a recording medium carrying a program to predict a protein or polypeptide (B) that interacts with a specific protein or polypeptide (A), comprising at least the following means a) to f):

- a) a means for inputting amino acid sequence information of the protein or polypeptide (A) and storing the information;
- b) a means for decomposing the above-mentioned information into a series of oligopeptides having a pre-determined length as sequence information, and a means for storing the sequence information consequently obtained;
- c) a means for storing an input protein database;
- d) a means for accessing the stored protein database and detecting a protein or polypeptide (C) having an amino acid sequence of said oligopeptide or a protein or polypeptide (D) having an amino acid sequence homologous to the amino acid sequence of said oligopeptide, and a means for storing and calculating a detected result;
- e) a means for carrying out local alignment between the protein or polypeptide (A) and the detected protein or polypeptide (C) or protein or polypeptide (D), and a means for storing and calculating a result; and
- f) a means for obtaining a resultant value of a frequency of an amino acid and/or a frequency of said oligopeptide from a protein database, followed by showing an index for predicting protein-protein interactions from the resultant value and a resultant value of said local alignment, and a means for storing and displaying the result and consequently detecting protein or polypeptide (B) which interacts with the protein or polypeptide (A).

In a further embodiment, the present invention relates to a recording medium comprising at least one of the following means g) to l) in addition to the means a) to f):

- g) a means for ranking strength of protein-protein interactions among detected proteins or polypeptides (B) based on the indexes calculated from a resultant value of local alignment and a resultant value of a frequency of an amino acid and/or a frequency of an oligopeptide in a protein database in the case that more than one protein or polypeptide (B) exist that are detected, and a means for storing and displaying the result;
- h) a means for displaying full-length of amino acid sequences of the protein or polypeptide (A) and the protein or polypeptide (B) that is detected, followed by indicating a location of partial sequence to be aligned in the full-length sequence in the case that amino acid partial sequences are aligned by local alignment between the protein or polypeptide (A) and the protein or polypeptide (B);
- i) a means for calculating a stereo structure model in the case that a stereo structure of the protein or polypeptide (A) or the protein or polypeptide (B) that is detected is known or in the case that homology modeling enable to make a stereo structure model, followed by displaying the structure of the amino acid partial sequences that are aligned by local alignment between the protein or polypeptide (A) and the protein or polypeptide (B) on the stereo structure;
- j) a means of classifying proteins in a protein database to narrow a searching area and storing the same;
- k) a means for serially inputting each protein in a protein database as the protein or polypeptide (A); and
- l) a means for storing a genome database.

In still another embodiment, the present invention relates to a device for predicting protein-protein interactions comprising the means that are carried by the above-mentioned recording medium.

In an additional embodiment, the present invention relates to a method for specifying proteins or polypeptides that interact with each other, which comprises

identifying a protein or polypeptide (B) that is predicted to interact with a specific protein or polypeptide (A) by the above-mentioned prediction method or prediction device, and then experimentally confirming the presence of the interaction between the protein or polypeptide (A) and the protein or polypeptide (B).

Furthermore, in another embodiment, the present invention relates to a protein or polypeptide that is specified by the above method.

In still another embodiment, the present invention relates to a method of screening for a compound that is capable of controlling the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B) utilizing the above-mentioned prediction method or prediction device.

In yet another embodiment, the present invention relates to a novel compound obtained by the screening method and a novel compound capable of controlling the interaction of the protein or polypeptide (A) with the protein or polypeptide (B) obtained by drug design based on information of the compound obtained.

In another embodiment, the present invention relates to an oligopeptide comprising amino acid sequence SEQ ID No: 1 which is capable of controlling the interaction of verotoxin 2 (VTII) with Bcl-2, or an oligopeptide that comprises an amino acid sequence homologous to the oligopeptide and is capable of controlling the interaction of VTII with Bcl-2, or a polypeptide that contains any of these oligopeptides and is capable of controlling the interaction of VTII with Bcl-2.

In addition, in one embodiment, the present invention relates to an agent against cell death comprising an oligopeptide comprising amino acid sequence SEQ ID NO: 1.

In still another embodiment, the present invention relates to a method of screening for a compound capable of controlling interaction of VTII with Bcl-2, wherein the method utilizes the above-mentioned oligopeptide and/or the above-mentioned polypeptide.

In yet another embodiment, the present invention relates to a method for determining a sequence of an oligonucleotide coding an oligopeptide involved in interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B) that is predicted to interact with the protein or polypeptide (A), wherein the method uses the above-mentioned prediction method or the above-mentioned prediction device.

In a further embodiment, the present invention relates to a series of combinations of human proteins, which are predicted to interact with each other, identified by the above-mentioned prediction method or the above-mentioned prediction device.

In addition, in an embodiment, the present invention relates to a method for selecting a combination of proteins having a protein-protein interaction that is related to a disease, wherein the method comprises selecting the combination based on the information of a known protein that is related to the disease from the above-mentioned series of combination of proteins.

Further in another embodiment, the present invention relates to a series of combinations of proteins having protein-protein interaction that are related to diseases, and which are obtained by the above-mentioned method.

In yet another embodiment, the present invention relates to a method of screening for a compound that controls the interaction of a certain combination and/or two proteins further selected from the series of combinations of proteins having a protein-protein interaction that are related to diseases obtained as mentioned above.

In a still further embodiment, the present invention relates to a compound identified by the method of screening for a compound which controls the interaction.

In yet another embodiment, the present invention relates to a method for predicting a processing site of a protein by predicting the protein-protein interaction of a specific protein with an enzyme cleaving said protein using the above-mentioned prediction method or device.

In addition, in one embodiment, the present invention relates to an amino acid sequence that contains a protein-processing site obtained by the above-mentioned prediction method for a protein-processing site, and/or an amino acid sequence that contains a partial sequence homologous to the processing site.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1a illustrates 20 amino acid residues from the amino terminal end of verotoxin 2.

Fig. 1b illustrates oligopeptides, each having an amino acid sequence length of 5 residues, which were obtained by decomposing the amino acid sequence consisting of said 20 residues as the sequence information using a program.

Fig. 1c illustrates oligopeptides, each having an amino acid sequence length of 6 residues, which were obtained by decomposing the amino acid sequence consisting of said 20 residues as the sequence information using a program.

Fig. 2 illustrates oligopeptides, each having an amino acid sequence length of 5 residues, which were obtained by decomposing the 13 residues from the amino terminal end of verotoxin 2 as the sequence information using a program, and human proteins comprising the amino acid sequence of the oligopeptides.

Fig. 3 illustrates the result of local alignment whereby oligopeptides that comprise a portion of verotoxin 2 (VTII) and human α -adrenergic receptor kinase 2 (ARK2) were obtained.

Fig. 4a illustrates the frequency of each amino acid in protein synthesis of *Escherichia coli*.

Fig. 4b illustrates the percentage that each amino acid is present in protein synthesis of *Escherichia coli*.

Fig. 5 is a simplified flow of means.

Fig. 6 illustrates amino acid sequences of oligopeptides derived from verotoxin 2 which are also present in human proteins that are related to cell death, and the corresponding human proteins.

Fig. 7 illustrates the result of local alignment whereby oligopeptides that comprise a portion of verotoxin 2 (VTII) and Bcl-2, were obtained.

Fig. 8 illustrates the result of local alignment whereby oligopeptides that comprise a portion of verotoxin 2 (VTII) and Bcl-xL, were obtained.

Fig. 9 illustrates the result of local alignment whereby oligopeptides that comprise a portion of verotoxin 2 (VTII) and MCL-1, were obtained.

Fig. 10a-b illustrate the result of local alignment whereby oligopeptides that comprise a portion of verotoxin 1 (VTI) and Bcl-2 (Fig. 10a) or Bcl-xL (Fig. 10b), were obtained.

Figs. 11a-b illustrate electrophoretic patterns showing the result of confirmational experiments using HepG2 cells and B10 cells showing that verotoxin 2 (VTII) and Bcl-2 interact with each other ((Fig 11a) and (Fig 11b) right), but that verotoxin 1 (VTI) and Bcl-2 do not interact with each other ((Fig. 11b) left). In the figures, Bcl-2 IPs and VTII IPs indicate that Bcl-2 and VTII were immunoprecipitated by anti-Bcl-2 antibody and anti-VTII antibody, respectively. Fig. 11a (left) illustrates the result of western blotting using anti-Bcl-2 antibody (Bcl-2 WB). Fig. 11a (right) illustrates the result of western blotting using anti-VTII antibody (VTII WB). Fig. 11b illustrates electrophoretic patterns showing the result of confirmation of the subcellular fraction of B10 cells that were treated with verotoxin 1 (VTI) (left) or verotoxin 2 (VTII) (right), where these proteins were detected using anti-VTI antibody and anti-VTII antibody, respectively.

Fig. 12 illustrates the sites which correspond to the local alignment of verotoxin 2 (VTII) and Bcl-2.

Fig. 13 illustrates, using a wire model, the portion that is homologous to the partial sequence of verotoxin 2 (VTII) on the stereo structure of Bcl-xL.

Fig. 14 illustrates, using a wire model, the portion that is homologous to the partial amino acid sequence of Bcl-xL on the stereo structure of verotoxin 2 that is constructed by homology modeling.

Fig. 15 illustrates that oligopeptide NWGRI which comprises a portion of verotoxin 2 (VTII) and Bcl-2, suppresses cell death induced by VTII in a dose dependent manner of NWGRI.

Figs. 16a-b illustrate the result of local alignment whereby oligopeptides that comprise a portion of human helper T cell surface protein CD4 and HIV-1 virus surface protein gp120, were obtained. Fig. 16a illustrates oligopeptides that comprise a portion of CD4 and gp120. Fig. 16b illustrates amino acid sequences of a region having a high local homology in CD4 and gp120.

Fig. 17 illustrates the result of local alignment whereby oligopeptides that comprise a portion of CED-4 (a cell death-related protein of nematode) and MAC-1 protein (which binds to CED-4), were obtained.

Fig. 18 illustrates the result of local alignment whereby oligopeptides that comprise a portion of amyloid precursor protein (APP) and BASE (an enzyme which cleaves the protein), were obtained.

Figs. 19a-b illustrate the result of local alignment whereby oligopeptides that comprise a portion of furin-precursor protein (furin-pre) and von Willebrand factor precursor protein (VWF-pre), were obtained. Fig. 19a illustrates oligopeptides that comprise a portion of both proteins. Fig. 19b illustrates an amino acid sequence of a region having a high local homology in both proteins.

Fig. 20 illustrates the result of local alignment whereby oligopeptides that comprise a portion of amyloid precursor protein (APP) and protein PC7 (which is

considered to be involved in the processing thereof), were obtained. In the figure, symbol “=” indicates a site that is predicted to be a cleavage site.

BRIEF DESCRIPTION OF REFERENCE NUMERALS

- a Means for inputting
- b Means for decomposing into a series of oligopeptides and storing the same
- c Means for storing
- d Means for searching and storing
- e Means for carrying out local alignment and storing
- f Frequency-calculating/memory-displaying means
- g Ranking/memory-displaying means
- h Location-displaying means
- i Stereo structure-calculating/memory-displaying means
- j Means for classifying proteins and storing
- k Sequentially inputting means
- l Means for storing
- m Keyboard
- n Controlling means
- o Outputting means

DETAILED DESCRIPTION OF THE INVENTION

The embodiments of the present invention will be described in more detail below as well as the principle and method of the present invention, a recording medium that carries a program for carrying out the method, a device that works the function, proteins and polypeptides that are obtained by the method and device. The following description is given only for illustration, and it not intended to limit the present invention.

Technical and scientific terms used in the specification have the meanings usually understood by one of ordinary skill in the art to which the present invention pertains, unless otherwise defined. Reference is made herein to various methodologies known to those of ordinary skill in the art. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties.

In one embodiment, the present invention relates to a method for predicting protein-protein interactions, which is based on the following idea: protein is composed of a sequence consisting of 20 kinds of amino acids, but these amino acids are not randomly placed. Therefore, it is considered that an oligopeptide that is a partial sequence of a protein has a role in a species of a living organism.

For example, an oligopeptide that is a part of a certain enzyme is considered to play a role in recognizing a substrate. In another protein, an oligopeptide that plays an important role in interacting with other proteins is considered to exist. In this way, it is necessary to consider the function or interaction of a protein from the oligopeptide level. In addition, from the viewpoint of frequency of the oligopeptide, the frequency of the appearance of certain oligopeptides in all of the proteins encoded by the genome in one organism is not even. Some oligopeptides frequently occur in various proteins; others do so only rarely. It is very likely that an oligopeptide that occurs with low frequency is an oligopeptide that is unique to each protein. Such an oligopeptide might determine the feature or function of the protein.

On the other hand, the fact that proteins interact with each other implies that the interacting proteins perform a function in cooperation with each other whereby the organism carries on its biotical activity. If it is assumed that one oligopeptide corresponds to one function, two proteins that interact with each other might have the same

oligopeptide or homologous oligopeptides. In addition, these two proteins might have homologous sequence structure in a part other than the oligopeptide that is the same.

As one of the techniques of similarity search for analyzing homology of two proteins, a method of comparison by aligning the primary structures of both proteins is known (Minoru Kanahisa “Introduction to genome informatics (in Japanese)” Kyoritsu Publishing Co., Ltd., 93-104, 1996). This sequence alignment includes ‘global alignment’ and ‘local alignment’. The ‘global alignment’ comprises aligning the entire sequences, and the ‘local alignment’ comprises locally aligning only homologous parts extracted from their sequences. In any alignment, the alignment is carried out so that the relation between/among two sequences or more can be as clear as possible. Many combinations exist in the alignment depending on the length of the sequence. Methods for carrying out combinatorial optimization include the dynamic programming method. The Smith-Waterman method (Smith, TF and Waterman, MS, J. Mol. Biol. 147, 195-197, 1981) that is based on the principle that dynamic programming gives an estimation function on the combinatorial optimization of sequences. A value of the estimation function, i.e., ‘homology score’ or ‘score’ permits estimating homology between two proteins. It can be estimated that the higher the score between proteins that were compared, the higher the homology between these proteins. As for the local alignment, it is carried out by setting a threshold to the score, followed by carrying out combinatorial optimization of partial sequences, when the combination of sequences is searched by dynamic programming (e.g., Gotoh, O., Pattern matching of biological sequences with limited storage, Comput. Appl. Biosci. 3, 17-20, 1987). The local alignment method permits searching the homologous structure in a part of the protein other than where the oligopeptide that is the same portion to two proteins is located.

Protein-protein interactions might have been conserved in the process of evolution. The case of verotoxin of *Escherichia coli* and Bcl-2, as described later, implies

that one function has been conserved in the protein-protein interaction beyond species, whence structurally similar amino acid sequences might exist. In addition, in the processing of amyloid precursor protein and von Willebrand Factor (VWF) precursor protein, as described later, the function of proteolysis might have been conserved in the protein-protein interaction, and structurally similar amino acid sequences might exist.

A method for predicting protein-protein interactions that is created based on the above idea may also permit predicting a network of functions that has been known only as a single function in the past and describing a new image of life based on the results that were predicted on a computer and on the overall relation of actions that is different from the image of life being reached by the accumulation of facts that have been obtained by the enumeration principle of molecular biology.

In addition, if the prediction for interactions is possible not in one organism but between two organisms, e.g., human beings and pathogenic microorganisms, the elucidation of the pathogenesis that has not been known so far might become possible.

Concretely, one embodiment of the above method for predicting protein-protein interactions is a method of extracting and predicting a counter-protein from a protein database and the like, which interacts with a protein that was obtained by genome analysis or cDNA analysis whose function is unknown or a protein whose function is known, wherein the method comprises, for example, the following steps 1-4:

In step 1, an amino acid sequence of a specific protein or polypeptide is decomposed into a series of oligopeptides having a pre-determined length as the sequence information. In step 2, proteins or polypeptides are determined which comprise each oligopeptide. In step 3, homology of partial structures between the proteins is estimated by local alignment. In step 4, each oligopeptide is further evaluated by a frequency of occurrence.

Each step will be described below more in detail:

Step 1:

The amino acid sequence of a specific protein or polypeptide (A), such as a protein or polypeptide that is obtained by genome analysis or cDNA analysis and whose function is not known or a protein or polypeptide whose function is known, is decomposed into oligopeptides as sequence information by shifting, serially, by one amino acid residue from the amino terminal end to the carboxyl end.

For example, Figs. 1a-b illustrate oligopeptides (Fig. 1b) having an amino acid length of 5 residues that were obtained by decomposing the first 20 residues (Fig. 1a) of verotoxin 2 (VTII) of *Escherichia coli* O157:H7 from its amino terminal end as sequence information. Amino acids and oligopeptides are given in their one-letter symbols hereafter.

When step 1 is carried out, the amino acid length of oligopeptide that is decomposed as the sequence information is 4 to 15 residues, preferably 4 to 8 residues. The longer the length of an oligopeptide, the greater the particularity of the oligopeptide, as shown in Example 2.

Step 2:

In step 2, a protein or polypeptide (C) comprising an amino acid sequence of an oligopeptide that was obtained by the decomposition in step 1 or a protein or polypeptide (D) having an amino acid sequence that is homologous to the oligopeptide is searched for in an amino acid sequence database of proteins or polypeptides. The number of detected proteins or polypeptides (C) or (D) can be large or can be one depending on the oligopeptide used.

For example, Fig. 2 illustrates the results of searching for proteins having 9 oligopeptides each consisting of 5 amino acids obtained by decomposing 13 residues of verotoxin 2 (VTII) of *Escherichia coli* O157:H7 from the amino terminal end, in a protein database (SWISS-PROT version 35). Verotoxin 2 causes food poisoning and/or renal

damage in human beings, so that human protein can be targets for searching for counter-proteins that interact with VTII. For example, such a search shows that a human protein comprising oligopeptide KCILF shown in Fig. 2 (second) is α -adrenergic receptor kinase 2 (ARK2 HUMAN).

Step 3:

Local alignment is carried out between the above protein or polypeptide (A) and the protein or polypeptide (C) or protein or polypeptide (D) that is obtained in the search in Step 2.

For example, Fig. 3 illustrates the result of local alignment between verotoxin 2 (VTII) and α -adrenergic receptor kinase 2 (ARK2).

Step 4:

If the result of the local alignment in step 3 shows any homology of partial sequence between the above protein or polypeptide (A) and the detected protein or polypeptide (C) and/or protein or polypeptide (D), the protein or polypeptide (C) and/or protein or polypeptide (D) are/is predicted to possibly be a protein or polypeptide (B) that interacts with protein or polypeptide (A). Moreover, the frequency of amino acid(s) and/or the frequency of oligopeptide that is present in both protein or polypeptide (A) and the detected protein or polypeptide (C) and/or protein or polypeptide (D) is calculated from a protein database, followed by evaluating the particularity of each oligopeptide in the protein database or in the genome of an organism having the protein or polypeptide (A) or the detected protein or polypeptide (C) and/or protein or polypeptide (D). If the particularity is high, the reliability of the prediction is evaluated to be high that the protein or polypeptide (C) and/or protein or polypeptide (D) are/is a protein or polypeptide (B) that interacts with the protein or polypeptide (A).

For example, an index of particularity of the above oligopeptide KCILF is calculated to be 1284.86×10^{-10} from the composition ratio (Fig. 4b) that is calculated

from the frequency of amino acid (Fig. 4a) in all of the proteins encoded by the genome of *Escherichia coli* shown in Fig. 4, so that the particularity is high. An oligopeptide consisting of 5 amino acids that is calculated to have low particularity in the *E. coli* genome is LLLLL, i.e., the particularity index is $136344.34 \times 10^{-10}$. An oligopeptide consisting of 5 amino acids that is calculated to have the highest particularity is CCCCC, with a particularity index of 2.208×10^{-10} , but the oligopeptide is not found in the *E. coli* genome. Therefore, the prediction that verotoxin 2 interacts with α -adrenergic receptor kinase 2 is evaluated to have high reliability from the value of the particularity index.

In order to confirm further the protein-protein interaction, the gene encoding the protein may be cloned for expression based on the information of the obtained proteins that interact with each other. For example, as described in the examples later, VTII and Bcl-2 were predicted to interact with each other by the above predicting method or a predicting device carrying a program for the predicting method. This could be confirmed by experiments wherein cells in which Bcl-2 is expressed and cells in which Bcl-2 is not expressed were treated with VTII, followed by co-immunoprecipitating with anti-Bcl-2 antibody and anti-VTII antibody. In addition, it is possible to specify the oligopeptide as an important interacting site, for example, by introducing a mutation by a well-known method into the amino acid sequence of an oligopeptide that is predicted to be an interacting site, followed by confirming that the interaction is lost. The method for confirming interactions experimentally is not limited to the above ones, but any of techniques that are applicable by those skilled in the art may be used.

In addition, if it is confirmed that an oligopeptide, which was predicted to be an interacting site, interrupts a protein-protein interaction and suppresses any function or action of the protein, then such an oligopeptide can be utilized as an agent for suppressing the action of the protein. For example, as described in an example below, oligopeptide NWGRI, which was predicted to be the interacting site for VTII and Bcl-2, suppresses

cell death induced by VTII and can be used as an agent against cell death. Such a low-molecular-weight compound can be utilized as pharmaceuticals, reagents, and the like.

Next, a recording medium and device that carry a program for the above method for predicting protein-protein interactions will be described. The above recording medium and device comprise at least the following means (a) to (f). Fig. 5 illustrates an example of the constitution.

Inputting means (a):

A means for inputting the amino acid sequence information concerning a specific protein or polypeptide (A) such as a protein or polypeptide, which was obtained by genome analysis or cDNA analysis, whose function is not known or protein or polypeptide whose function is known.

Means for decomposing into a series of oligopeptides and storing the same (b):

A means for decomposing the amino acid sequence information that was input by inputting means (a) into a series of oligopeptides having a pre-determined length as sequence information by shifting, serially, by one amino acid residue from the amino terminal end to the carboxyl end, and storing the result.

Storing means (c):

A means for storing a database that was input in concerning with a protein or polypeptide.

Searching/storing means (d):

A means for accessing a database concerning a protein or polypeptide that is stored in storing means (c), followed by searching for a protein or polypeptide (C) comprising the amino acid sequence of the above oligopeptide or a protein or polypeptide (D) comprising an amino acid sequence that is homologous to the amino acid sequence of the above oligopeptide, and storing the result.

Carrying out local alignment/storing means (e):

A means for carrying out local alignment between the above protein or polypeptide (A) and the detected protein or polypeptide (C) and/or (D), and storing the result.

Frequency-calculating/memory-displaying means (f):

A means for calculating an index for predicting protein-protein interactions from the result of the above local alignment and the result obtained after calculating a frequency of an amino acid and/or a frequency of an oligopeptide in a peptide or polypeptide database, and storing and displaying the result.

In addition, in the above program for predicting protein-protein interaction, it is also possible to comprise the following means (g) to (l) in an appropriate combination:

Ranking/memory-displaying means (g):

A means having a function of ranking proteins or polypeptides (B), when more than one protein or polypeptide (B) is detected, by using the result of the local alignment and the result of the calculation of a frequency of an amino acid and/or a frequency of an oligopeptide from a protein database as indexes and a function of storing/displaying the result.

Location-indicating means (h):

A means for displaying full-length amino acid sequences of the protein or polypeptide (A) and the protein or polypeptide (B) followed by indicating a location of partial sequence to be aligned in the full-length sequences in the case that amino acid partial sequences are aligned between the protein or polypeptide (A) and the detected protein or polypeptide (B).

Stereo structure-calculating/memory-displaying means (i):

A means for calculating a stereo structure model followed by displaying the structure of the amino acid partial sequences that are aligned between the protein or

polypeptide (A) and the protein or polypeptide (B) on the stereo structure in the case that a stereo structure of the protein or polypeptide (A) or the protein or polypeptide (B) that is detected is known or in the case that a stereo structure model can be constructed by homology modeling.

Protein-classifying/storing means (j):

A means having a function of classifying proteins or polypeptides in a protein or polypeptide database by feature, function, and/or origin to narrow a searching area followed by storing them.

Sequentially inputting means (k):

A means for sequentially inputting each protein or polypeptide in a protein or polypeptide database as the protein or polypeptide (A).

Storing means (l):

A means having a function of storing a genome database.

The above means are carried on an appropriate medium.

As one embodiment of use, each of these means may be provided as a device containing a recording medium selectively carrying it as a program. A device for predicting protein-protein interactions is operated as described below (see Fig. 5).

A specific protein or polypeptide (A), such as a protein or peptide which was input by an inputting means (a) and whose function is unknown or known, is decomposed by means for decomposing into a series of oligopeptides and storing the same (b) into a series of oligopeptides having a pre-determined length as sequence information, and the oligopeptides are stored. In this case, the protein or polypeptide (A) is sequentially input by a sequentially inputting means (k) from a means (c) that stores a database that was input concerning the protein or polypeptide, when desired. A search is carried out through a means (c) for the amino acid sequence of the above-stored oligopeptide by a

searching/storing means (d), and a protein or polypeptide (C) comprising the amino acid sequence of the oligopeptide or a protein or polypeptide (D) comprising an amino acid sequence that is homologous to the amino acid sequence of the oligopeptide is detected and stored. When searching, it is also possible to classify proteins or polypeptides in a database to narrow the searching area by a protein-classifying/storing means (j), followed by searching within the resultant area. The above protein or polypeptide (A) and the detected protein or polypeptide (C) or (D) are subjected to local alignment by a locally aligning/memory-displaying means (e), and the result is stored. Next, by a frequency-calculating/storing means (f), the frequency of an amino acid and/or the frequency of the oligopeptide are/is calculated from a database that was stored on a means (c), and an index for predicting protein-protein interactions is calculated from the result and the above-obtained result of the local alignment and stored. Then, those obtained are displayed on the screen of the device, which are the protein or polypeptide (C) or (D) that are predicted to interact with the above protein or polypeptide (A), an amino acid sequence of an oligopeptide that is the same in these proteins, a frequency of the oligopeptide, indexes for predicting protein-protein interactions, and so on. Displayed results permits giving a protein or polypeptide (B) that has interaction with the above protein or polypeptide (A) based on the indexes for predicting protein-protein interactions. In addition, concerning the above protein or polypeptide (B), it is also possible to display the functional information of a protein that is stored on a means (c) and the gene information from a means (l) equipped when desired that stores a genome database. When more than one protein or polypeptide (B) is detected, a ranking/memory-displaying means (g) permits ranking the protein or polypeptide (B) in order of the particularity to interact with the above protein or polypeptide (A). It is also possible to indicate by a location-indicating means (h) which part of the full-length amino acid sequences of the above protein or polypeptide (A) and the detected protein or polypeptide (B) is the partial

amino acid sequence that is aligned between the protein or polypeptide (A) and the protein or polypeptide (B). In addition, it is also possible to display a stereo structure of the protein or polypeptide (A) and the protein or polypeptide (B), as well as the part that is aligned between the protein or polypeptide (A) and the protein or polypeptide (B) by a stereo structure-calculating/memory-displaying means (i). This device may be equipped with keyboard (m), controlling means (n), outputting means (o), as shown also in Fig. 5, and so on as well as these means (a) – (l).

The above method for predicting protein-protein interactions or the above prediction device can further be used for screening for a novel compound that controls the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B). The above method of screening for a novel compound that controls the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B) is carried out based on the information of the amino acid sequence of a key oligopeptide. An amino acid sequence of a selected oligopeptide, an amino acid sequence of an oligopeptide homologous thereto, or a polypeptide comprising the amino acid sequence or the homologous amino acid sequence per se can be capable of controlling the interaction of the protein or polypeptide (A) with the protein or polypeptide (B). For example, in the case that the protein or polypeptide (B) having a receptor function to the protein or polypeptide (A) is in existence, it is likely that an oligopeptide that is screened by the above technique is antagonistic to the interaction of the protein or polypeptide (A) with the protein or polypeptide (B). For example, in the case that the protein or polypeptide (A) is activated by the interaction with the protein or polypeptide (B), it is likely that an oligopeptide that is screened by the above technique has a function as an agonist.

Concretely, as described in detail in the examples below, it was experimentally confirmed that an oligopeptide NWGRI described in SEQ ID NO: 1, which comprises a portion of VTII and Bcl-2, that were predicted and experimentally confirmed to interact

with each other by the present invention, interrupts complex formation due to the interaction of VTII with Bcl-2, and suppresses cell death induced by VTII. Therefore, NWGRI oligopeptide can be used as a medicament for controlling a disease related to cell death induced by VTII, for example as a medicament for treating a disease caused by *Escherichia coli* O157 expressing VTII, more concretely as an agent against cell death. Moreover, an oligopeptide having an amino acid sequence homologous to the amino acid sequence and capable of controlling the interaction of VTII with Bcl-2, or a polypeptide comprising the amino acid sequence or an amino acid sequence homologous to the amino acid sequence and capable of controlling the interaction of VTII with Bcl-2, can also be used as a medicament for controlling a disease related to cell death induced by VTII. In addition, a novel compound capable of controlling the interaction of VTII with Bcl-2 can be obtained utilizing these oligopeptides and polypeptides by the drug design method or by applying of a known screening method.

In this way, a novel compound, which is obtained by drug design based on the information of an oligopeptide that is obtained by the above screening method according to an embodiment of the present invention, is capable of controlling the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B). Namely, to predict interaction of the above protein or polypeptide (A) with the above protein or polypeptide (B) permits one to make a derivative of the oligopeptide obtained by the above screening method and a low-molecular-weight compound having a structure homologous to the oligopeptide by a well-known drug design technique.

The above prediction method is also very useful for a method for determining the sequence of the oligonucleotide coding an oligopeptide involved in interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B). Applying well-known methods such as substitution, deletion, addition, insertion, or induced mutation based on this information permits one to obtain a useful oligonucleotide. The

obtained oligonucleotide can be used for obtaining a compound for controlling the interaction of protein or polypeptide (A) with protein or polypeptide (B) on a gene level. For example, it is utilized for making an antisense oligonucleotide to interrupt the protein-protein interaction. In addition, the obtained oligonucleotide can be used for diagnosing a disease that is related to the protein-protein interaction.

In another embodiment, the present invention relates to a series of combinations of human proteins that are predicted to have protein-protein interactions, which are predicted by the above method or device for predicting protein-protein interaction. The series of combination of proteins can be provided as a catalogue or as a database. A series of combination of proteins which interact with each other that are involved in a disease can be obtained by selecting ones having protein-protein interactions that are related to a disease based on the information of known proteins that can be related to the disease from the series of combination of proteins having protein-protein interactions. These can be provided as a catalogue or as a database. These combinations of proteins are useful as a medicament for treating or preventing diseases or as ways to obtain medicaments. For example, a compound capable of controlling interaction of two proteins can be obtained by screening using a well-known screening method and by utilizing a combination of proteins that is obtained.

In the case that among combinations consisting of two proteins having a protein-protein interaction, one protein is an enzyme capable of processing protein and cleaves the other protein, the processing site of the protein that is cleaved can be predicted by the above method or device for predicting protein-protein interactions.

For example, as shown in an example below, prediction could be accomplished on the subject of the interaction of an amyloid precursor protein with an enzyme that is involved in its processing, and on the subject of the interaction of von Willebrand factor precursor protein with an enzyme furin that is involved in its processing. Namely, the

above method or device for predicting protein-protein interactions permits predicting a cleavage site when a precursor protein is cleaved to act as a mature protein. In this way, a hitherto-unknown enzyme having a protein-processing action related to a disease and a protein that is cleaved by the enzyme can be obtained by predicting protein-protein interactions.

Examples

Although advantages, features, and possible applications of the present invention are described below in greater detail with reference to exemplary embodiments, the present invention is not limited to the following examples. In addition, although SWISS-PROT version 35 was used as a protein database in the following examples, other protein databases or the like can also be used.

Example 1

Figs. 1a-c illustrate oligopeptides that were decomposed from the first 20 residues (Fig. 1a) of verotoxin 2 (VTII) of *Escherichia coli* O157:H7 from the amino terminal end, where the oligopeptides have an amino acid sequence length of 5 residues (Fig. 1b) as an example of step 1. Fig. 1c illustrates oligopeptides that were decomposed from the first 20 residues of verotoxin 2 (VTII) from the amino terminal end, where the oligopeptides have an amino acid sequence length of 6 residues.

Example 2

In step 4 of the above method for predicting protein-protein interactions, values are used as an index for predicting the interaction of proteins or polypeptides. The values are calculated from the frequency of an amino acid in a protein or polypeptide database and the frequency of an oligopeptide in the protein or polypeptide database. By way of

example, the particularity of oligopeptides is calculated from the frequency of the amino acid in all of the proteins encoded by the genome of *Escherichia coli* shown in Figs. 4a-b. The percentage 'Ai' of each amino acid 'ai' can be calculated to be as shown in Fig. 4b from the frequency of occurrence (Fig. 4a) of the 20 kinds of amino acids in all of the proteins encoded by the genome of *Escherichia coli*.

The particularity of oligopeptide $a_1a_2a_3a_4a_5$ is calculated to be $A_1 \times A_2 \times A_3 \times A_4 \times A_5$. For example, in the case of oligopeptide KCILF, it is calculated to be $4.406610 \times 1.170608 \times 6.004305 \times 10.639652 \times 3.898962 \times 10^{-10}$. The particularity of oligopeptide LLLLL is calculated to be $136344.34 \times 10^{-10}$, and the particularity of oligopeptide CCCCC is calculated to be 2.20×10^{-10} .

The smaller the value, the greater the particularity of the oligopeptide. The oligopeptide that has the highest particularity among those having an amino acid sequence length of 5 residues is oligopeptide CCCCC, but this oligopeptide does not occur in any of the proteins encoded by the genome of *Escherichia coli*. In contrast, the oligopeptide that has the lowest particularity is oligopeptide LLLLL.

When a protein or polypeptide (A) is decomposed into oligopeptides in step 1, the longer the oligopeptide, the greater the particularity of the oligopeptide.

Example 3

In step 4 of the above method for predicting protein-protein interactions, as an index for predicting the interaction of proteins or polypeptides, the result of the local alignment is used. Here is mentioned an example in which scores of the alignment of a partial sequence by Gotoh's method (Gotoh, O., Pattern matching of biological sequences with limited storage, Comput. Appl. Biosci. 3, 17-20, 1987) are used. In the following examples, when the score was 25.0 or higher, it was judged that the partial amino acid

sequences are aligned (homologous) between a protein or polypeptide (A) and another protein or polypeptide (B).

The “m” amino acid partial sequences are premised to be aligned between a protein or polypeptide (A) and another protein or polypeptide (B) with their scores being $S_i (1 \leq i \leq m)$ and the amino acid length of protein or polypeptide (B) being LB. The index for predicting the interaction of protein or polypeptide (A) with protein or polypeptide (B) calculated from the result of local alignment is defined as the sum $(S_i)/LB$. It is predicted that the higher the index, the stronger the interaction.

Example 4

Prediction of interaction of VTII with Bcl-2

Verotoxin 2 (VTII) of *Escherichia coli* O157:H7 causes food poisoning and renal damage in human beings, but the mechanism of action is not well-known (Sandvig, K., et al., Exp. Med. Biol. 412, 225-232, 1997; Paton, JC., and Paton, AW. Clin. Microbiol. Rev. 11, 450-479, 1998). This protein is a toxic protein. Therefore, human proteins relating to cell death serve as candidates of proteins interacting with this protein. Thus, human proteins that may interact with this protein were searched, specifically for human proteins relating to cell death, in protein database SWISS-PROT version 35, and an example is given below showing that they actually interact with each other.

Among oligopeptides having an amino acid sequence length of 5 residues of verotoxin 2, those found to be contained in a human protein relating to cell death were the following four, i.e., LCLLL, QRVAA, EFSGN, and NWGRI, in SWISS-PROT version 35 (see Fig. 6, where the human proteins are shown by using protein IDs of SWISS-PROT version 35). Values of particularity for these oligopeptides were calculated from the amino acid frequencies in all of the proteins encoded by the genome of *Escherichia coli* shown in Figs. 4a-b, i.e., the particularity of LCLLL was

15001.03×10^{-10} ; that of QRVAA was 15584.55×10^{-10} ; that of EFSGN was 3801.65×10^{-10} ; that of NWGRI was 1479.85×10^{-10} . It was found that NWGRI has the highest particularity among these four oligopeptides.

Oligopeptide NWGRI comprises a portion of verotoxin 2 and each of three human proteins, i.e., Bcl-2, Bcl-xL, and MCL-1. Local alignment between verotoxin 2 (VTII) and each of Bcl-2, Bcl-xL, and MCL-1 revealed partial homology in their amino acid sequence, as shown in Figs. 7, 8, and 9. Then, the sum of the scores of the local alignment was divided by the length of each protein to give index as described in Example 3, and shown below.

$$\text{Bcl-2} \quad (30.0 + 27.0 + 25.0) / 239 = 0.343$$

$$\text{Bcl-xL} \quad (30.0 + 29.0 + 27.0) / 233 = 0.369$$

$$\text{MCL-1} \quad (34.0 + 30.0 + 28.0 + 26.0) / 350 = 0.337$$

Among these three proteins, Bcl-2 and Bcl-xL constitutes the same family. Based on the index calculated from the local alignment by the above method, the prediction is that Bcl-2 and Bcl-xL have higher interaction with verotoxin 2 among Bcl-2, Bcl-xL and MCL-1.

Verotoxin 1 (VTI) is one of the verotoxins produced by *Escherichia coli* O157:H7, and is an isoform of verotoxin 2. The toxicity of verotoxin 1 is weaker than that of verotoxin 2, with the former being about one fiftieth the latter (Tesh, VL., et al., 1993, Infect. Immun. 61, 3392-3402). In protein database SWISS-PROT version 35, a human protein that contains an oligopeptide having an amino acid length of 5 residues that comprises a portion of verotoxin 1 and is related to cell death is P2X1_HUMAN, the oligopeptide being SSTLG. However, the particularity of oligopeptide SSTLG, which is calculated to be 14385.63×10^{-10} from the amino acid frequencies in all of the proteins encoded by the genome of *Escherichia coli* shown in Figs. 4a-b, is lower than that of NWGRI, by about one tenth.

In verotoxin 1, the oligopeptide NWGRL that corresponds to oligopeptide NWGRI having an amino acid length of 5 residues in verotoxin 2 reveals a particularity of 2622.30×10^{-10} , calculated from Figs. 4a-b, that is lower than that of NWGRI. Both Bcl-2 and Bcl-xL contain oligopeptide NWGR having an amino acid sequence length of 4 residues that comprises a portion of verotoxin 1. Comparison between the particularity of NWGRI and that of NWGRL permits prediction that both Bcl-2 and Bcl-xL interact more strongly with verotoxin 2 than with verotoxin 1. In addition, the indexes obtained by the calculation from the result (Fig. 10) of the local alignment between verotoxin 1 and Bcl-2 or Bcl-xL are $(27.0 + 26.0)/239 = 0.222$ and $26.0/233 = 0.112$, respectively (there is no homologous amino acid partial sequence other than the NWGR part). Consequently, it is predicted that the interaction of verotoxin 1 with Bcl-2 or Bcl-xL is considerably weaker than the interaction of verotoxin 2 with Bcl-2 or Bcl-xL.

Example 5

Experimental confirmation of prediction of interaction of VTII with Bcl-2

In Example 4, the reliability of prediction that verotoxin 2 interacts with human Bcl-2 or Bcl-xL was predicted to be high. Based on the result of this prediction, it was experimentally confirmed that verotoxin 2 actually interacts with Bcl-2 (Fig. 11a and Fig. 11b (right)). Specifically, human hepatic cancer cell HepG2 (essentially not expressing the Bcl-2 gene) and B10 cells prepared by transducing a Bcl-2-expressing vector into HepG2 so as to express Bcl-2, were treated with verotoxin 2 (VTII), and then co-immunoprecipitation was conventionally carried out using anti-Bcl-2 antibody (Bcl-2 IPs) and anti-VTII antibody (VTII IPs).

Fig. 11a (left) illustrates the result of the western blotting analysis using anti-Bcl-2 antibody after co-immunoprecipitation; Fig. 11a (right) illustrates the result of the western blotting analysis using anti-VTII antibody. It was confirmed from these

results that a VTII/Bcl-2 complex was co-immunoprecipitated in the B10 cells, i.e., these two proteins interact with each other. Moreover, B10 cells were treated with verotoxin 1 (VTI) or verotoxin 2 (VTII) to examine in which subcellular fraction these proteins were detected using anti-VTI antibody and anti-VTII antibody. Bcl-2 in mitochondria plays a very important role in cell death. Verotoxin 2 (VTII) was detected also in a mitochondria fraction (Fig. 11b (right)).

On the other hand, verotoxin 1 was not detected in the mitochondria fraction. Thus, it was proved experimentally that verotoxin 1 does not have a strong interaction with mitochondria Bcl-2. The result is shown in Fig. 11b (left).

Example 6

Fig. 12 illustrates an example wherein the full-length amino acid sequences of verotoxin 2 (VTII) and Bcl-2 were displayed so as to show the locations of the partial sequences aligned in the full-length sequences.

Example 7

The stereo structure of Bcl-xL is known, with the structure being registered in PDB, that is a protein stereo structure database. Based on the result of the local alignment of Fig. 8, partial amino acid sequences homologous to those of verotoxin 2 in the stereo structure of Bcl-xL are shown with bold lines in Fig. 13.

Example 8

Verotoxin 2 is believed to cleave a part of ribosomal RNA so as to stop protein synthesis, thereby exerting its toxicity. The stereo structure of protein 'ricin' that cleaves a part of ribosomal RNA is registered in PDB, that is a protein stereo structure database. Based on the structure, homology modeling of verotoxin 2 was carried out. Based on the

result of the local alignment of Fig. 8, the amino acid partial sequences homologous to those of Bcl-xL is shown in the stereo structure model with bold lines in Fig. 14.

Example 9

Suppression by NWGRI of cell death induction by VTII

Next, it was experimentally confirmed that oligopeptide NWGRI (SEQ ID NO: 1), which was found in Example 4 and comprises a portion of VTII and Bcl-2, can control the interaction of VTII with Bcl-2. First of all, the complex formation was examined using an extract of the Bcl-2-expressing B10 cells used in Example 5 and biotinylated VTII in the presence of oligopeptide, and then analyzed by Far Western blotting analysis. Oligopeptide NWGRI interrupted the complex formation of VTII and Bcl-2 in a dose dependent manner.

In addition, B10 cells were pretreated with oligopeptide NWGRI at 0, 10, 50, 100 μ M and were treated with VTII at 10 ng/ml for 24 hr, and the induction of cell death by apoptosis was assayed. A total of about 5,000 nuclei was dyed with Hoechst 33342/PI (Propidium iodide) according to the conventional method, and the ratio of nuclei that showed apoptosis is shown in Fig. 15. As shown in the figure, about 85% of cells caused apoptotic cell death by the treatment with only VTII, while the induction of apoptotic cell death was suppressed by pretreatment with oligopeptide NWGRI in a dose dependent manner. Thus, it was confirmed that oligopeptide NWGRI, which comprises a portion of VTII and Bcl-2, interrupts the interaction of VTII with Bcl-2 so as to inhibit the complex formation of these proteins and suppresses cell death induction by VTII thereupon.

Example 10

CD4/gp120HIV-1

Human AIDS virus HIV-1 infects helper T cells. An important first step to infecting these cells is that protein gp120 on the viral surface binds to surface protein CD4 of helper T cells. In this example, it was examined if the binding of gp120 and CD4 can be predicted by the above prediction method.

Protein CD4 was decomposed into oligopeptides having an amino acid sequence length of 5 residues, and proteins having the amino acid sequence of the oligopeptide derived from CD4 were serially searched in a protein database, and gp120 was extracted as a protein that contains oligopeptide SLWDQ (Fig. 16a). Oligopeptide SLWDQ exists only in protein CD4 as a human protein in SWISS-PROT version 35, i.e., the frequency in human proteins is 1 and the particularity is very high. Moreover, besides this oligopeptide, a locally homologous region exists (Fig. 16b). It is known that amino acid residue arginine (Arg) next to oligopeptide SLWDQ in the N-terminal side and 67-SFLTKGP-73 play important roles when CD4 binds to gp120 (Kwong, PD., et al., Nature, vol. 398, 648-659, 1998). It is also known that a few amino acid residues next to the homologous region (289-KTIIVQLNETVKINCIRPNNKT-310) shown in Fig. 16b in the N-terminal side is one of the regions playing an important role when CD4 is recognized by gp120 (Kwong, PD., et al., Nature, vol. 398, 648-659, 1998). Therefore, even if the binding between gp120 and CD4 is not known, it can be predicted by the above prediction method.

Example 11

CED-4/MAC-1

Nematode *Caenorhabditis elegans* is the first multicellular organism whose entire genome information was elucidated. One example concerning *C. elegans* is described here. Protein CED-4 plays a central role in the control of programmed cell death. MAC-1 was found to be a protein that binds to CED-4 and suppresses cell death

(Wu et al., Development, vol. 126, 9, 2021-2031, 1999). Therefore, oligopeptides that comprise a portion of these two proteins were examined so as to verify the present invention, although the binding between MAC-1 and CED-4 is known. As a result, it was found that MAC-1 and CED-4 contain the same oligopeptide FPSVE having an amino acid sequence length of 5 residues, and the present invention was verified. The index of this oligopeptide, calculated from a frequency of amino acids in the genome of *C. elegans*, was 5.436. Moreover, as illustrated in Fig. 17, there are many homologous regions between these two proteins, whereby the binding of these proteins was strongly suggested (top sequence, CED-4; bottom sequence, MAC-1).

Example 12

APP/BASE

APP (amyloid precursor protein), which is one of the proteins causing Alzheimer's disease, gives rise to amyloid upon being cleaved at two sites. An enzyme (BASE, bata secretase) that cleaves the site on the amino terminal side of the two cleavage sites was recently discovered (VASSAR et al., Science, 286(5440), 735-741, 1999). Cleavage of APP by BASE indicates the presence of the interaction of these two proteins. To verify the present invention, oligopeptides that comprise a portion of these two proteins were examined. APP and BASE have homologous oligopeptides WYFDV and WYYEV having an amino acid sequence length of 5 residues that comprise a portion of each of them. The oligopeptide WYFDV exists only in protein APP as a human protein in SWISS-PROT version 35. A human protein comprising WYYEV is not registered yet. Both oligopeptides have high particularity. This result verified the present invention. Fig. 18 illustrates the regions homologous between the two proteins (top sequence, APP; bottom sequence, BASE).

Example 13

Furin and von Willebrand factor

Furin is an intracellular serine protease, and is related to the secretion system pathway, such as von Willebrand factor (VWF), albumin, and complement C3. An example of the interaction of furin with VWF is mentioned here. VWF is cleaved from a precursor protein by furin to act as a mature protein. Cleavage of the VWF precursor protein by furin requires the interaction of these two proteins. Moreover, furin *per se* becomes a mature protein from a precursor protein of furin by being cleaved to act as a protease. Therefore, to verify the present invention, an oligopeptide that comprises a portion of furin precursor protein and VWF precursor protein was examined. The two proteins comprising the same oligopeptide HCPPG, at positions 613-617 of furin precursor protein and at positions 1176-1180 of VWF precursor protein (Fig. 19a). Both locations are within the regions of the mature proteins. The oligopeptide HCPPG comprises a portion of only furin precursor protein and VWF precursor protein as human proteins in SWISS-PROT version 35, and has high particularity from the viewpoint of frequency. VWF precursor protein is cleaved by furin at the site between the 763rd amino acid residue and the 764th amino acid residue. Local alignment between furin precursor protein and VWF precursor protein reveals that the region near the site of VWF precursor protein cleaved by furin has a partial region homologous to furin precursor protein (Fig. 19b). Thus, even if a novel protein was presumed to be a protease by the motif of the active part, and a counterpart protein as well as the cleavage site in the counter protein was not known, the present invention permits predicting the counterpart protein, as well as the cleavage site in the counter protein.

Example 14

APP and PC7

APP alpha is a peptide formed by cleavage of amyloid precursor protein (APP) at a site different from the two cleavage sites to form amyloid. It was recently found that PC7 (proprotein convertase subtilisin/kexin type 7) is involved in cleavage for forming APP alpha (Lopez-Perez E et al., J. Neurochem., vol. 73, 5, 2056-2062, 1999). Examination of the oligopeptide that comprises a portion of the two proteins APP and PC7 revealed that APP and PC7 have homologous oligopeptides DSDPSG and DSDPNG having an amino acid sequence length of 6 residues that comprise a portion of both of them. The oligopeptide DSDPSG exists only in protein APP as a human protein in SWISS-PROT version 35. A human protein comprising DSDPNG is not registered yet. Both oligopeptides have very high particularity. This result verified the present invention. Fig. 20 illustrates regions homologous between the two proteins.

Between K and L of 687-KLVFFAEDVGS-697 of APP in Fig. 20 is the cleavage site to form APP alpha, and Fig. 20 illustrates that a partial sequence (359-RMPFYAEECAS-369) homologous to this exists in PC7. Namely, this example shows that the present invention permits predicting a protein involved in cleaving another protein.

INDUSTRIAL APPLICABILITY

As described above in detail, the present invention permits predicting, by using a protein database, a counterpart protein that interacts with a protein having an unknown function that is obtained by genome analysis or cDNA analysis or a protein having a known function. Namely, the protein-protein interaction in one organism whose genome information was elucidated can be predicted on a computer using a protein database based on genome analysis and cDNA analysis that have been recently enhanced. If the prediction on a computer becomes possible, information concerning proteins that were predicted to interact with each other based on the prediction on a computer can be easily

obtained without adopting a risky technique wherein the result depends on a cDNA library used, such as the two-hybrid method. The prediction became possible, so that it becomes possible to easily predict the sequence of an oligopeptide involved in the interaction, and to design a novel compound capable of controlling protein-protein interactions based on the information. The present invention makes elucidating protein-protein interactions efficient, and can be widely utilized in various fields including biochemistry, molecular biology, pharmaceutical development, agriculture, and biotechnology. Especially, in the development of pharmaceuticals, the present invention permits predicting the mechanism of disease that has not so far been known, and gives a possibility of creating novel pharmaceuticals.

What is claimed is:

1. A method for predicting a protein or polypeptide (B) that interacts with a specific protein or polypeptide (A), wherein the method is characterized by comprising:
 - 1) decomposing the amino acid sequence of protein or polypeptide (A) into a series of oligopeptides having a pre-determined length as sequence information;
 - 2) searching, within a database of protein or polypeptide amino acid sequences, for a protein or polypeptide (C) comprising an amino acid sequence for each member of the series or for a protein or polypeptide (D) comprising an amino acid sequence homologous to an amino acid sequence for each member of the series;
 - 3) carrying out local amino acid sequence alignment between said protein or polypeptide (A) and the detected protein or polypeptide (C) or detected protein or polypeptide (D); and
 - 4) predicting whether the detected protein or polypeptide (C) and/or protein or polypeptide (D) is a protein or polypeptide (B) that interacts with the protein or polypeptide (A) based on the results of the local amino acid sequence alignment and a value calculated from a frequency of amino acids and/or a frequency of said oligopeptides in said amino acid sequence database.
2. The method according to claim 1, wherein the oligopeptide is 4-15 amino acids in length.
3. A recording medium carrying a program to predict a protein or polypeptide (B) that interacts with a specific protein or polypeptide (A), wherein the recording medium is characterized by comprising at least the following means a)~f):
 - a) a means for inputting amino acid sequence information of the protein or polypeptide (A) and storing the information;

- b) a means for decomposing said information into a series of oligopeptides having a pre-determined length as sequence information, and a means for storing the sequence information consequently obtained;
 - c) a means for storing an input protein database;
 - d) a means for accessing the stored protein database and detecting a protein or polypeptide (C) having an amino acid sequence of said oligopeptide or a protein or polypeptide (D) having an amino acid sequence homologous to the amino acid sequence of said oligopeptide, and a means for storing and calculating a detected result;
 - e) a means for carrying out local alignment between said protein or polypeptide (A) and the detected protein or polypeptide (C) or protein or polypeptide (D), and a means for storing and calculating a result; and
 - f) a means for obtaining a resultant value of a frequency of an amino acid and/or a frequency of said oligopeptide from a protein database, followed by showing an index for predicting protein-protein interactions from the resultant value and a resultant value of said local alignment, and a means for storing and displaying the result and consequently detecting protein or polypeptide (B) which interacts with the protein or polypeptide (A).
4. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3:
- g) a means for ranking strength of protein-protein interactions among detected proteins or polypeptides (B) based on the indexes calculated from a resultant value of local alignment and a resultant value of a frequency of an amino acid and/or a frequency of an oligopeptide in a protein database in the case that more than one protein or polypeptide (B) exist that are detected, and a means for storing and displaying the result.

5. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3 or 4:
 - h) a means for displaying full-length of amino acid sequences of said protein or polypeptide (A) and said protein or polypeptide (B) that is detected, followed by indicating a location of partial sequence to be aligned in the full-length sequence in the case that amino acid partial sequences are aligned by local alignment between the protein or polypeptide (A) and the protein or polypeptide (B).
6. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3, 4 or 5:
 - i) a means for calculating a stereo structure model in the case that a stereo structure of said protein or polypeptide (A) or said protein or polypeptide (B) that is detected is known or in the case that homology modeling enables to make a stereo structure model, followed by displaying the structure of the amino acid partial sequences that are aligned by local alignment between the protein or polypeptide (A) and the protein or polypeptide (B) on the stereo structure.
7. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3, 4, 5 or 6:
 - j) a means of classifying and storing proteins in a protein database to narrow a searching area.
8. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3, 4, 5, 6 or 7:
 - k) a means for serially inputting each protein in a protein database as said protein or polypeptide (A).
9. A recording medium characterized by comprising at least the following means in addition to the means according to claim 3, 4, 5, 6, 7 or 8:
 - l) a means for storing a genome database.

10. A device for predicting protein-protein interactions which comprises the means that are carried by the recording medium according to claim 3, 4, 5, 6, 7, 8, or 9.
11. A method for specifying proteins or polypeptides that interact with each other, which comprises identifying a protein or polypeptide (B) that is predicted to interact with a specific protein or polypeptide (A) using the method according to claim 1 or 2, and then experimentally confirming the presence of the interaction between the protein or polypeptide (A) and the protein or polypeptide (B).
12. A method for specifying proteins or polypeptides that interact with each other, which comprises identifying a protein or polypeptide (B) that is predicted to interact with a specific protein or polypeptide (A) using the device according to claim 10, and then experimentally confirming the presence of the interaction between the protein or polypeptide (A) and the protein or polypeptide (B).
13. A protein or polypeptide that is specified by the method according to claim 11 or 12.
14. A method of screening for a compound that controls the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B), wherein the method utilizes the method according to claim 1 or 2.
15. A method of screening for a compound that controls the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B), wherein the method uses the device according to claim 10.
16. A novel compound obtained by the screening method according to claim 14 or 15.
17. A novel compound capable of controlling the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B), which is obtained by drug design based on information of a compound obtained by the screening method according to claim 14 or 15.
18. An oligopeptide comprising the amino acid sequence of SEQ ID No: 1, which is capable of controlling the interaction of verotoxin 2 (VTII) with Bcl-2.

19. An agent against cell death comprising an oligopeptide comprising the amino acid sequence of SEQ ID NO: 1.
20. An oligopeptide comprising an amino acid sequence homologous to the amino acid sequence of SEQ ID No: 1, which is capable of controlling the interaction of VTII with Bcl-2.
21. A polypeptide comprising the amino acid sequence of the oligopeptide according to claim 18 or 20, which is capable of controlling the interaction of VTII with Bcl-2.
22. A method of screening for a compound capable of controlling the interaction of VTII with Bcl-2, wherein the method utilizes the oligopeptide according to claim 18 or 20 and/or the polypeptide according to claim 21.
23. A method for determining the nucleotide sequence of an oligonucleotide coding an oligopeptide which is involved in the interaction of a specific protein or polypeptide (A) with a protein or polypeptide (B), wherein the method uses the prediction method according to claim 1 or 2 or the prediction device according to claim 10.
24. A series of combinations of human proteins that are predicted to interact with each other, obtained by the method according to claim 1 or 2 or the device according to claim 10.
25. A method for selecting a combination of proteins which interact with each other, wherein said interaction is related to a disease, wherein the method comprises selecting the combination based on the information of a known protein that can be related to the disease from the series of combinations according to claim 24.
26. A series of combinations of proteins which interact with each other, wherein said interaction is related to a disease, wherein each member of said series is selected according to the method of claim 25.

27. A method of screening for a compound which controls the interaction of a combination of proteins and/or two proteins selected from the series of combinations according to claim 26.
28. A compound obtained by the method according to claim 27.
29. A method for predicting a processing site of a protein comprising predicting the interaction of a specific protein with an enzyme cleaving said protein using the method according to claim 1 or 2 or the device according to claim 10.
30. A polypeptide comprising an amino acid sequence that is predicted to contain a protein-processing site obtained by the method according to claim 29 and/or to contain a partial sequence homologous to the processing site.

ABSTRACT

The present invention relates a method for predicting a protein or polypeptide (B) that interacts with a specific protein or polypeptide (A), wherein the method is characterized by comprising:

- 1) decomposing the amino acid sequence of protein or polypeptide (A) into a series of oligopeptides having a pre-determined length as sequence information;
 - 2) searching, within a database of protein or polypeptide amino acid sequences, for a protein or polypeptide (C) comprising an amino acid sequence for each member of the series or for a protein or polypeptide (D) comprising an amino acid sequence homologous to an amino acid sequence for each member of the series;
 - 3) carrying out local amino acid sequence alignment between said protein or polypeptide (A) and the detected protein or polypeptide (C) or detected protein or polypeptide (D); and
 - 4) predicting whether the detected protein or polypeptide (C) and/or protein or polypeptide (D) is a protein or polypeptide (B) that interacts with the protein or polypeptide (A) based on the results of the local amino acid sequence alignment and a value calculated from a frequency of amino acids and/or a frequency of said oligopeptides in said amino acid sequence database; and to
- a recording medium for carrying out the above method, a device comprising the recording medium, and proteins obtained thereby.

1/17
DRAWINGS

Fig.1

(a)

MKCILFKWVLCLLLGFSSVS

(b)

MKCIL
KCILF
CILFK
ILFKW
LFKWV
FKWVL
KWVLC
WVLCL
VLCLL
LCLLL
CLLLG
LLLGF
LLGFS
LGFS
GFSSV
FSSVS

(c)

MKCILF
KCILFK
CILFKW
ILFKWV
LFKWVL
FKWVLC
KWVLCL
WVLCLL
VLCLLL
LCLLLG
CLLLGF
LLLGFS
LLGFSS
LGFSV
GFSSVS

Fig.2

MKCIL

none

KCILF

BETA-ADRENERGIC RECEPTOR KINASE 2 (ARK2_HUMAN)

CILFK

WHITE PROTEIN HOMOLOG (WHIT_HUMAN)

ILFKW

none

LFKWV

none

FKWVL

none

KWVLC

none

WVLCL

none

VLCLL

GROWTH ARREST AND DNA-DAMAGE-INDUCIBLE PROTEIN GADD45
(GA45_HUMAN)

Fig.3

>Score = 32.0

ARK2 618 KCILFR 623

VTII 2 KCILFK 7

>Score = 27.0

ARK2 651 FKEAQRLLRRA 661

**= ** =*=*

VTII 193 FRQIQREFRQA 203

>Score = 26.0

ARK2 299 TEIILGLEHV 308

*** ****

VTII 44 TEISTPLEHI 53

>Score = 26.0

ARK2 379 CMLFK 383

VTII 3 CILFK 7

>Score = 25.0

ARK2 12 SYLMAMEKSKATPAARASKRI 32

*** ** * * = *** *

VTII 160 SYLALMEFSGNTMTRDASRAV 180

Fig. 4

(a)

L: 144615
A: 128924
G: 100162
V: 95941
I: 81611
S: 79179
E: 78102
R: 75293
T: 73489
D: 69829
Q: 60214
P: 60182
K: 59895
N: 53727
F: 52995
M: 38745
Y: 38721
H: 30912
W: 20761
C: 15911

(b)

L: 10.639652%
A: 9.485230%
G: 7.369144%
V: 7.058596%
I: 6.004305%
S: 5.825378%
E: 5.746140%
R: 5.539476%
T: 5.406752%
D: 5.137477%
Q: 4.430080%
P: 4.427726%
K: 4.406610%
N: 3.952817%
F: 3.898962%
M: 2.850557%
Y: 2.848791%
H: 2.274266%
W: 1.527434%
C: 1.170608%

Fig.5

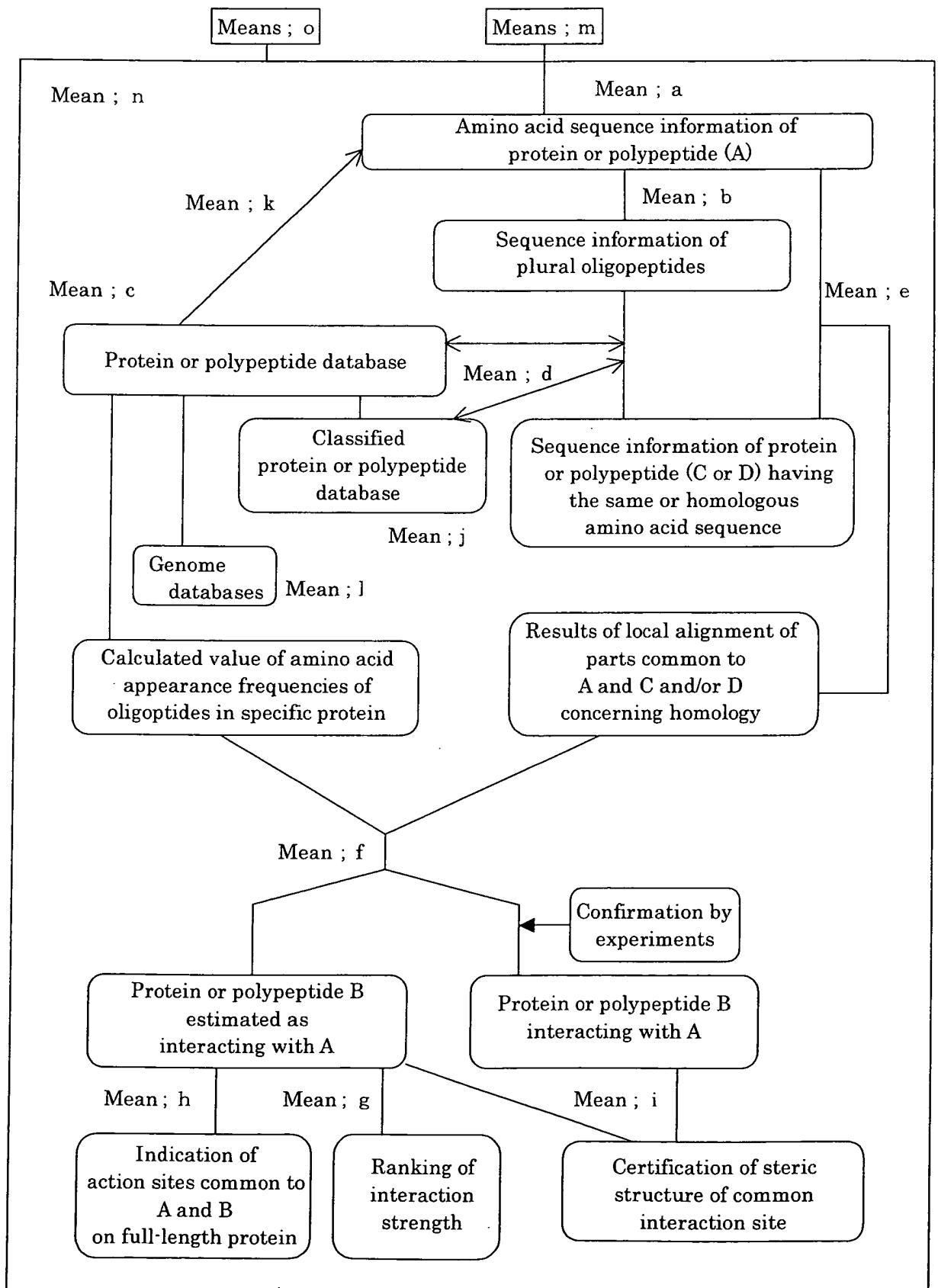


Fig.6

```

LCLLL    FASA_HUMAN
QRVAA    TRAI_HUMAN
EFSGN    DAPK_HUMAN
NWGRI    BCL2(BC2A)_HUMAN, BCLX_HUMAN, MCL1_HUMAN

```

Fig.7

>Score = 30.0

```

Bcl-2  143 NWGRI 147
      *****

```

```

VTII   223 NWGRI 227

```

>Score = 27.0

```

Bcl-2  129 RFATVVEELFR 139
      ** ** * _*

```

```

VTII   182 RFVTVTAEALR 192

```

>Score = 25.0

```

Bcl-2  156 VMCVESVNREMSPLVDNIA 174
      *   ** *** *_*=

```

```

VTII   36 VSSLNSIRTEISTPLEHIS 54

```

Fig.8

>Score = 30.0

Bcl-xL 136 NWGRI 140

VTII 223 NWGRI 227

>Score = 29.0

Bcl-xL 101 YRRAFSDLTSQLHITPG 117

__** _****

VTII 200 FRQALSETAPVYTMTPG 216

>Score = 27.0

Bcl-xL 5 NRELVVDF 12

VTII 22 SREFTIDF 29

Fig.9

>Score = 34.0

```

Mc1-1  326 GGIRNVLLAFAGVAGVGAG 344
          * * *** * * ** *
VTII    225 GRISNVLPYRGEDGVRVG 243

```

>Score = 30.0

```

Mc1-1  260 NWGRI 264
          *****
VTII    223 NWGRI 227

```

>Score = 28.0

```

Mc1-1  275 AKHLKTINQES 285
          ** *_**_**
VTII    268 ARSVRAVNEES 278

```

>Score = 26.0

```

Mc1-1  134 LGKRPAVLPLLELVGESGNTSTDGS 159
          = *== * = * *_ *** = * *
VTII    152 ISRHSLVSSYLALMEFSGNTMTRDAS 177

```

Fig.10

(a)

>Score = 27.0

Bcl-2 129 RFATVVEELFR 139

** ** * =*

VTI 182 RFVTVTAEALR 192

>Score = 26.0

Bcl-2 143 NWGRI 147

****=

VTI 224 NWGRL 228

(b)

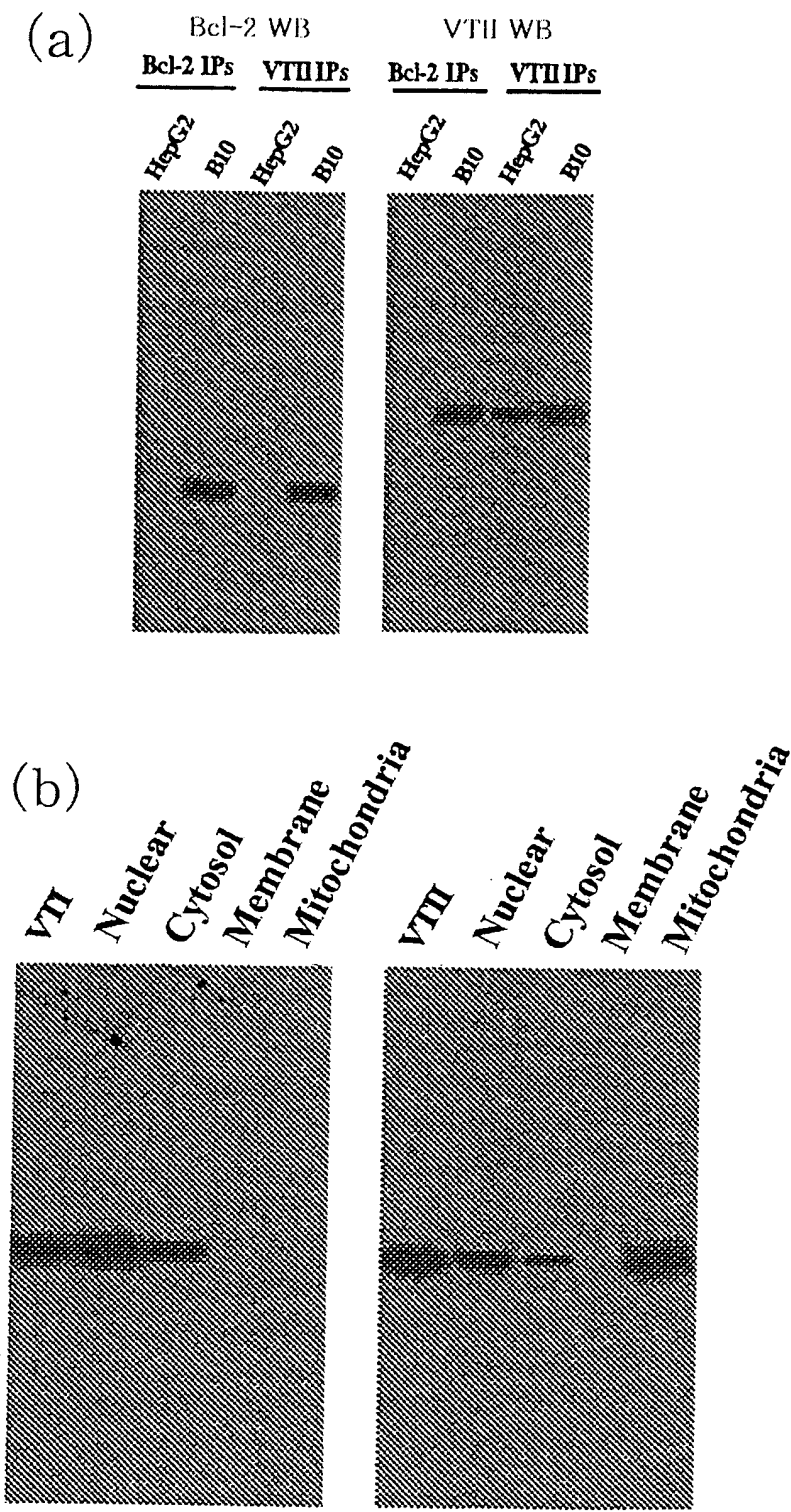
>Score = 26.0

Bcl-xL 136 NWGRI 140

****=

VTI 224 NWGRL 228

Fig.11



(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2001年9月13日 (13.09.2001)

PCT

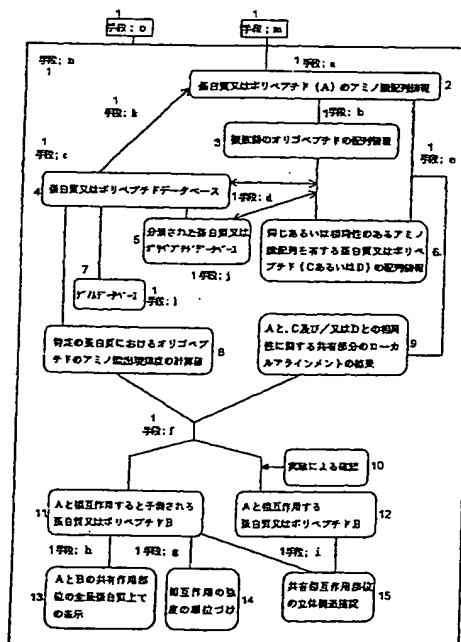
(10) 国際公開番号
WO 01/67299 A1

- (51) 国際特許分類⁷: G06F 17/30, C07K
7/06, 14/435, G01N 33/68
- (21) 国際出願番号: PCT/JP01/01846
- (22) 国際出願日: 2001年3月9日 (09.03.2001)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願2000-72485 2000年3月10日 (10.03.2000) JP
- (71) 出願人 (米国を除く全ての指定国について): 第一製薬株式会社 (DAIICHI PHARMACEUTICAL CO., LTD.)
[JP/JP]; 〒103-8234 東京都中央区日本橋3丁目14番10号 Tokyo (JP). 富士通株式会社 (FUJITSU LIMITED)
- (72) 発明者; および
(75) 発明者/出願人 (米国についてののみ): 土居 洋文 (DOI, Hirofumi) [JP/JP]; 〒273-0865 千葉県船橋市夏見五丁目29番4-515号 Chiba (JP). 鈴木 敦 (SUZUKI, Atsushi) [JP/JP]; 〒134-8630 東京都江戸川区北葛西一丁目16番13号 第一製薬株式会社 東京研究開発センター内 Tokyo (JP).
- (74) 代理人: 庄司 隆, 外 (SHOJI, Takashi et al.); 〒101-0032 東京都千代田区岩本町3丁目2番10号 SN岩本町ビル6階 Tokyo (JP).
- (81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT,

[続葉有]

(54) Title: METHOD OF ANTICIPATING INTERACTION BETWEEN PROTEINS

(54) 発明の名称: 蛋白質間相互作用予測方法



- 1...AMINO ACID SEQUENCE DATA OF PROTEIN OR POLYPEPTIDE (A)
- 2...AMINO ACID SEQUENCE DATA OF PLURAL OLIGOPEPTIDES
- 3...PROTEIN OR POLYPEPTIDE DATA BASE
- 4...CLASSIFIED PROTEIN OR POLYPEPTIDE DATA BASE
- 5...SEQUENCE DATA OF PROTEIN OR POLYPEPTIDE (C OR D) HAVING THE SAME OR HOMOLOGOUS AMINO ACID SEQUENCE
- 6...SEQUENCE DATA OF PROTEIN OR POLYPEPTIDE (C OR D) HAVING THE SAME OR HOMOLOGOUS AMINO ACID SEQUENCE
- 7...GENOME DATA BASE
- 8...CALCULATED AMINO ACID APPEARANCE FREQUENCIES OF OLIGOPEPTIDES IN SPECIFIC PROTEIN
- 9...RESULTS OF LOCAL ALIGNMENT OF PARTS COMMON TO A AND C AND/OR D CONCERNING HOWELOC
- 10...CONFIRMATION BY EXPERIMENT
- 11...PROTEIN OR POLYPEPTIDE B ANTICIPATED AS INTERACTING WITH A
- 12...PROTEIN OR POLYPEPTIDE B INTERACTING WITH A
- 13...INDICATION OF ACTION SITES COMMON TO A AND B ON FULL-LENGTH PROTEIN
- 14...INTERACTION STRENGTH RANKING
- 15...CERTIFICATION OF STERIC STRUCTURE OF COMMON INTERACTION SITES

(57) Abstract: A method of anticipating an interaction between proteins characterized by comprising: (1) digesting the amino acid sequence of a protein A to give oligopeptides of certain length; (2) searching for a protein C having the above-described oligopeptides or a protein D having oligopeptides homologous with the above-described oligopeptides from a protein data base; (3) performing local alignment between the above-described protein A and the protein C or D thus detected; and (4) anticipating that the detected protein C or D is

[続葉有]



LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT,
TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

添付公開書類:
— 国際調査報告書

- (84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

a protein B having an interaction with the protein A by using a value calculated from the amino acid frequencies or oligopeptide frequencies based on the results of the local alignment and the protein data base; a recording medium carrying the above anticipation program; an anticipation apparatus carrying this recording medium; and proteins obtained therefrom.

(57) 要約:

- (1) 蛋白質Aのアミノ酸配列をある長さのオリゴペプチドに分解し、
- (2) 前記オリゴペプチドを持った蛋白質Cを、又は前記オリゴペプチドと相同なオリゴペプチドを持った蛋白質Dを、蛋白質データベース中で検索し、
- (3) 前記蛋白質Aと検出された蛋白質C又はDとの間でローカルアライメントを行い、
- (4) ローカルアライメントの結果及び蛋白質データベースからアミノ酸の頻度あるいはオリゴペプチドの頻度から計算される数値を指標とすることによって、検索された蛋白質C又はDが蛋白質Aと相互作用を有する蛋白質Bであると予測することを特徴とする蛋白質間相互作用の予測方法、該予測プログラムを担持する記録媒体、及び該記録媒体を担持する予測装置、並びにこれらにより得られた蛋白質に関する。

明細書

蛋白質間相互作用予測方法

技術分野

- 5 本発明は、蛋白質間の相互作用の予測に関するものであり、そのための方法と装置、並びに該方法及び該装置により得られた蛋白質に関する。

背景技術

- 多くの蛋白質は、他の蛋白質又は同じ蛋白質と相互作用しながらその機能を果たしている。その蛋白質間の相互作用が明らかになることは、
10 医薬品の開発や農業での品種改良などにおいて重要である。特に、病原微生物を含む種々の生物のゲノム解析やcDNA解析の進展にともない、新たに見出された機能未知の遺伝子やそれによってコードされる蛋白質の数が急速に増加している。蛋白質間相互作用の解明がなされれば、それら機能未知の蛋白質の機能予測を行うことができるであろう。
15

- 蛋白質間の相互作用を解明する目的で、ある蛋白質についてそれと相互作用する蛋白質を検索するために採られる従来の方法は、two-hybrid法 (Field, S. The two-hybrid system to detect protein-protein interaction. METHODS: A Companion to Meth. Enzymol., 5, 116-124, 1993) が主なものとして知られている。しかし、two-hybrid法は探索的実験手段であって、操作が煩雑で時間がかかる上、得ることのできる蛋白質の数は意外と少ない。また、この方法は使用するcDNAライブラ
20 リーの質に結果が依存する、という欠点がある。すなわち、ある蛋白質についてそれと相互作用する蛋白質の遺伝子配列が、使用するcDNAライブラリーに入っていないというリスクを伴う。
25

一方、ゲノム解析やcDNA解析に基づく蛋白質データベースが充実してきたことから、細胞内の複合体を直接MALDI-TOFマスペクトルにかけ、そのアミノ酸配列の断片情報を蛋白質データベース中で探索する、という方法も採られている(Yates, JR 3rd, J. Mass. Spectrom, 33, 1-19, 1998; Hunphery-Smith, I., et al., Electrophoresis, 18, 1217-1242; Kaufmann, R., 1995, J. Biotechnol., 41, 155-175, 1997)。しかし、この方法では複合体(コンプレックス; complex)の構成蛋白質の情報は得られるが、蛋白質間の相互作用の情報は得られないため、
10 どの蛋白質とどの蛋白質が相互作用しているかについては実験的に確認する必要がある。

発明の開示

- 15 本発明の1つの態様は、特定の蛋白質もしくはポリペプチド(A)と相互作用する蛋白質もしくはポリペプチド(B)を予測する方法であって、
- (1) Aのアミノ酸配列をある長さの複数個のオリゴペプチドに配列情報として分解し、
- 20 (2) 前記各オリゴペプチドのアミノ酸配列を持った蛋白質もしくはポリペプチド(C)を、又は前記オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を持った蛋白質もしくはポリペプチド(D)を、蛋白質もしくはポリペプチドのアミノ酸配列のデータベース中で検索し、
- (3) 前記Aと検出されたC又はDとの間でアミノ酸配列についてのローカルアライメントを行い、
- 25 (4) ローカルアライメントの結果、及び、蛋白質もしくはポリペプチドのアミノ酸配列のデータベースにおけるアミノ酸の頻度及び/又はオ

オリゴペプチドの頻度から計算される数値を指標として、検出された C 及び／又は D を、A と相互作用する蛋白質もしくはポリペプチド (B) であると評価することを特徴とする蛋白質間相互作用の予測方法である。

本発明の 1 つの態様は、前記予測方法において、オリゴペプチドのアミノ酸配列の長さが、アミノ酸 4 ～ 15 個である予測方法でありうる。

また本発明の 1 つの態様は、特定の蛋白質もしくはポリペプチド (A) と相互作用する蛋白質もしくはポリペプチド (B) を予測する方法を実施するプログラムを担持する記録媒体であって、少なくとも以下の (a) から (f) の手段を備えたことを特徴とする蛋白質間相互作用の予測プログラムを担持する記録媒体である；

- (a) A のアミノ酸配列情報を入力し、記憶する手段と、
- (b) この情報のある長さの複数個のオリゴペプチドに配列情報として分解する手段と、その結果の配列情報を記憶する手段と、
- 15 (c) 入力された蛋白質データベースを記憶する手段と、
- (d) 該記憶された蛋白質データベースにアクセスし、前記オリゴペプチドのアミノ酸配列を持った蛋白質もしくはポリペプチド (C) を、又は前記オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を持った蛋白質もしくはポリペプチド (D) を、検出する手段と、その検出結果を
- 20 記憶、計算する手段と、
- (e) 前記 A と検出された C 又は D との間でローカルアライメントを行う手段と、その結果を記憶、計算する手段と、
- (f) 前記ローカルアライメントの結果値、並びに、蛋白質データベースからアミノ酸の頻度及び／又はオリゴペプチドの頻度の結果値を得て、
- 25 その結果値から、蛋白質間相互作用の予測のための指標を提示する手段と、その結果を記憶し表示し、A と相互作用する B を検出する手段。

本発明の 1 つの態様は、前記 (a) ～ (f) の手段を担持する記録媒

体に、下記 (g) ~ (l) の手段から選ばれる少なくとも一つの手段を備えた記録媒体である；

(g) 検出される B が複数ある場合、ローカルアライメントの結果値、並びに蛋白質データベースにおけるアミノ酸の頻度及び／又はオリゴペプチドの頻度の結果値から計算される指標に基づき、該検出された複数の B の間での蛋白質間相互作用の強さの順位付けを行う手段と、その結果を記憶、表示する手段、

(h) ローカルアライメントにより、前記 A と前記検出された B との間でアミノ酸部分配列間でアライメントされる場合、当該 A と当該 B のアミノ酸配列の全長を表示し、アライメントされる部分配列が全長上でどこにあるかを表示する手段、

(i) 前記 A もしくは前記検出された B の立体構造が既知の場合、又はホモロジーモデリングで立体構造モデルが作れる場合、その立体構造モデルを計算し、その立体構造上で、ローカルアライメントにより当該 A と当該 B の間でアライメントされるアミノ酸部分配列の構造を表示する手段、

(j) 検索範囲を絞るために蛋白質データベース中の蛋白質を分類し、記憶する手段、

(k) 蛋白質データベース中の各蛋白質を前記 A として順次入力する手段、

(l) ゲノムデータベースを記憶する手段。

本発明の 1 つの態様は、前記の記録媒体が担持する手段を備えた蛋白質間相互作用の予測装置である。

本発明の 1 つの態様は、前記の予測方法又は前記の予測装置により特定の蛋白質又はポリペプチド (A) と相互作用すると予測された蛋白質

又はポリペプチド（B）を得、ついで実験的にAとBとの間に相互作用があることを確認する、相互作用を有する蛋白質又はポリペプチドの特定方法である。

さらに、本発明の1つの態様は、この特定方法で特定された蛋白質又は
5 ポリペプチドである。

本発明の1つの態様は、前記の予測方法又は前記の予測装置を利用する、特定の蛋白質又はポリペプチド（A）と当該Aと相互作用すると予測された蛋白質又はポリペプチド（B）との蛋白質間相互作用を調整する機能を有する化合物を選別する方法である。

10 さらに、本発明の1つの態様は、この選別方法によって得られる新規化合物、及び得られた化合物の情報を基にドラッグデザインして得ることができるAとBとの蛋白質間相互作用の調整機能を有する新規化合物である。

15 本発明の1つの態様は、本発明によって得られた、ペロ毒素2（VT II）とBcl-2との相互作用を調整する機能を有する配列表の配列番号1に示したアミノ酸配列を有するオリゴペプチド、又は該オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を有するオリゴペプチドであってVT IIとBcl-2との相互作用を調整する機能を有するオリゴペ
20 プチド、又はこれらのオリゴペプチドのいずれかを含むポリペプチドであってVT IIとBcl-2との相互作用を調整する機能を有するポリペプチドである。

また、本発明の1つの態様は、配列表の配列番号1に示すアミノ酸配列を有するオリゴペプチドを含有する抗細胞死剤でありうる。

25

本発明の1つの態様は、前記オリゴペプチド及び／又は前記ポリペプチドを利用する、VT IIとBcl-2との相互作用を調整する機能を有

する化合物の選別方法である。

本発明の1つの態様は、前記の予測方法又は前記の予測装置を用いることを特徴とする、特定の蛋白質又はポリペプチド（A）と当該Aと相
5 互作用すると予測された蛋白質又はポリペプチド（B）との蛋白質間相互作用に係るオリゴペプチドをコードするオリゴヌクレオチドの配列の決定方法である。

本発明の1つの態様は、前記の予測方法又は前記の予測装置により求
10 められた、蛋白質間相互作用があると予測されるヒト由来蛋白質の組み合わせ群である。また、本発明の1つの態様は、この組み合わせ群から、疾病に關与する可能性がある既知蛋白質の情報を基にして選択を行う、疾病に關与する蛋白質間相互作用を有する蛋白質の組み合わせの選択方法である。さらに、本発明の1つの態様は、この方法で得られた疾病に
15 關与する蛋白質間相互作用を有する蛋白質の組み合わせ群でありうる。

本発明の1つの態様は、前記得られた疾病に關与する蛋白質間相互作用を有する蛋白質の組み合わせ群からさらに選ばれるある組み合わせ及び／又は二つの蛋白質の相互作用を調整する化合物を選別する方法であ
20 る。さらに、本発明の1つの態様は、この相互作用を調整する化合物を選別する方法で得られた化合物である。

本発明の1つの態様は、前記の予測方法又は前記の予測装置を用いて、特定の蛋白質とこの蛋白質を切断する酵素との蛋白質間相互作用を予測
25 することにより行う、蛋白質のプロセシング部位を予測する方法である。

さらに、本発明の1つの態様は、前記蛋白質のプロセシング部位を予測する方法で得られた蛋白質プロセシング部位及び／又は該プロセシ

グ部位と相同な部分配列を含むと予測されるアミノ酸配列を有するポリペプチドである。

図面の簡単な説明

- 5 図1の(a)はペロ毒素2由来のアミノ末端からの20残基を、(b)と(c)はそれぞれ該20残基からなるアミノ酸配列を基にプログラム上で配列情報として分解されて得られたアミノ酸長5個と6個のオリゴペプチドを示す。

図2は、ペロ毒素2由来のアミノ末端からの13残基の配列を基にプログラム上で配列情報として分解されて得られたアミノ酸長5個のオリ
10 ゴペプチドと該オリゴペプチドのアミノ酸配列を有するヒトの蛋白質を示す。

図3は、ペロ毒素2とヒトの β アドレナリン受容体キナーゼ2 (ARK2) とが共有しているオリゴペプチドをローカルアライメントにより
15 得た結果を示す。

図4の(a)は大腸菌ゲノムにおける各アミノ酸の頻度を、(b)は組成比率を示す。

図5は、手段の構成を示す図である。

図6は、ペロ毒素2由来のオリゴペプチドであって、ヒトの細胞死に
20 関係した蛋白質が有するオリゴペプチドのアミノ酸配列と、該アミノ酸配列を有するヒトの細胞死に関係した蛋白質を示す。

図7は、ペロ毒素2 (VTII) とBcl-2 とが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す。

図8は、ペロ毒素2 (VTII) とBcl-xL とが共有しているオリ
25 ゴペプチドをローカルアライメントにより得た結果を示す。

図9は、ペロ毒素2 (VTII) とMCL-1 とが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す。

図10は、ベロ毒素1 (VTI) と、Bc1-2 (a) 又はBc1-xL (b) とが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す。

図11は、ベロ毒素1 (VTI) とBc1-2は相互作用せず [(b) の左図]、ベロ毒素2 (VTII) とBc1-2は相互作用すること [(a) 及び (b) の右図) をHepG2細胞及びB10細胞を用いて実験的に確認した結果を示す電気泳動図である。図中、Bc1-2 IPs及びVTII IPsは、それぞれ抗Bc1-2抗体及び抗VTII抗体を用いて免疫沈降したことを表す。また、(a) の左図は抗Bc1-2抗体を用いてウエスタンブロットしたもの (Bc1-2 WB)、(a) の右図は抗VTII抗体を用いてウエスタンブロット (VTII WB) したものである。また、(b) はB10細胞にベロ毒素1 (VTI) (左図) 又はベロ毒素2 (VTII) (右図) を作用させ、細胞内のどのフラクションからこれらの蛋白質が検出されるかを、それぞれ抗VTI抗体及び抗VTII抗体を用いて確認した結果を示す電気泳動図である。

図12は、ベロ毒素2 (VTII) とBc1-2とのローカルアライメントがどの位置で対応しているかを示す。

図13は、Bc1-xLの立体構造上でベロ毒素2 (VTII) と部分配列が相同な部分をワイヤモデルで示す。

図14は、ホモロジーモデリングされたベロ毒素2の立体構造上でBc1-xLと部分アミノ酸配列が相同な部分をワイヤモデルで示す。

図15は、ベロ毒素2 (VTII) とBc1-2とが共有しているオリゴペプチドであるNWGR Iが、VTIIによる細胞死誘導をNWGR Iの濃度依存的に抑制することを示す。

図16は、ヒトヘルパーT細胞表面蛋白質CD4とHIV1ウイルス表面蛋白質fp120とが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す図面である。(a) はCD4とgp120

が共有するオリゴペプチドを示す。(b)はCD4とgp120で局所的に相同性の高い領域のアミノ酸配列を示す。

図17は、線虫の細胞死関連蛋白質であるCED-4とCED-4に結合するMAC-1蛋白質とが共有しているオリゴペプチドをローカル
5 アライメントにより得た結果を示す。

図18は、アミロイドプレカーサープロテイン(APP)とこれを切断する酵素であるBASEとが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す。

図19は、フューリンプレカーサー蛋白質(furin-pre)と
10 フォンビルプラント因子プレカーサー蛋白質(VWF-pre)とが共有しているオリゴペプチドをローカルアライメントにより得た結果を示す図面である。(a)は両蛋白質が共有しているオリゴペプチドを示す。

(b)は両蛋白質で局所的に相同性の高い領域のアミノ酸配列を示す。

図20は、アミロイドプレカーサープロテイン(APP)とこのプロセッシングに関与すると考えられている蛋白質PC7とが共有している
15 オリゴペプチドをローカルアライメントにより得た結果を示す図面である。図中、＝は切断部位として予測できる箇所を示している。

符号の簡単な説明

- 20 a 入力手段
- b オリゴペプチド分解／記憶手段
- c 記憶手段
- d 検索／記憶手段
- e ローカルアライメント実行／記憶手段
- 25 f 頻度計算／記憶表示手段
- g 順位付け／記憶表示手段
- h 位置表示手段

- i 立体構造計算／記憶表示手段
- j 蛋白質分類／記憶手段
- k 逐次入力手段
- l 記憶手段
- 5 m キーボード
- n 制御手段
- o 出力手段

発明を実施するための最良の形態

- 10 以下、本発明の思想、方法、該方法に関するプログラムを担持する記録媒体、及びその機能を発揮する装置、並びに該方法及び該装置により得られた蛋白質又はポリペプチドについて、発明の実施の形態をさらに詳しく説明する。以下の詳細な説明は例示であり、説明のためのものに過ぎず、本発明を何ら限定するものではない。
- 15 また、本明細書中で使用されている技術的及び科学的用語は、別途定義されていない限り、本発明の属する技術分野において通常の知識を有する者により普通に理解される意味を持つ。本明細書中では当業者に既知の種々の方法が参照されている。そのような引用されている公知の方法を開示する刊行物などの資料はそれらを引用することにより本明細書
- 20 中にそれらの全体が完全に記載されているものとして導入される。

本発明の1つの態様である蛋白質間相互作用予測方法は、以下のとおりの考えに立脚してなされたものである。すなわち、蛋白質は20種類のアミノ酸残基の配列によって構成されているが、その並びはランダム

25 ではない。したがって、任意の生物種において、蛋白質の部分配列であるオリゴペプチドが生命活動における何らかの役割を持っていると考えられる。

例えば、ある酵素の一部であるオリゴペプチドは基質を認識する役目を担っていると考えられる。また、別の蛋白質では他の蛋白質と相互作用するのに重要な役目を果たしているオリゴペプチドが存在すると考えられる。このようにある蛋白質の機能又は相互作用はオリゴペプチドのレベルから考察することが必要である。さらには、オリゴペプチドの頻度という観点からすれば、一つの生物において、その生物のゲノムでコードされる全蛋白質中にあるオリゴペプチドが出現する頻度は均一ではない。種々の蛋白質に高頻度で出現するオリゴペプチドや、まれにしか出現しないものがある。このうち、低頻度で出現するオリゴペプチドは、個々の蛋白質に特異的に存在するオリゴペプチドである可能性が高く、その蛋白質の特性や機能を決定していると考えることができる。

一方で、蛋白質が相互作用するということは、生命活動を営む上で相互作用する蛋白質間で協同で一つの機能を遂行することである。一つの機能に対してその役割を果たすために一つのオリゴペプチドが対応していると考えれば、相互作用する二つの蛋白質は同じオリゴペプチド又は相同なオリゴペプチドを共有していることになる。また、共有しているオリゴペプチド以外の部分においても二つの蛋白質は相同の配列構造をしている部分があると考えられる。

二つの蛋白質の相同性を解析するための類似性検索法の一手法として、両蛋白質の一次配列を並べて（アライメント；alignment）比較する方法が知られている（金久 實、ゲノム情報への招待、共立出版株式会社、93-104、1996）。この配列アライメントにはグローバルアライメントとローカルアライメントの区別があり、グローバルアライメントは配列全体を並べることで、ローカルアライメントは類似性の高い部分のみをそれぞれの配列から取り出して局所的に並べることである。いずれのアライメントにおいても、比較する二つの、又はそれ以上の数の、配列の関係ができるだけ明らかになるような並べ方をする。こ

の並べ方には、配列の長さに依存する多くの組み合わせが存在する。この組み合わせの最適化を計る方法としてダイナミックプログラミング法がある。ダイナミックプログラミングの原理に基づいたSmith-Waterman法 (Smith, TF and Waterman, M S, J. Mol. Biol. 147, 195-197, 1981) により配列の組み合わせの最適化についての評価関数を得、その値であるホモロジースコア (以下、単にスコアと呼ぶ) により、二つのタンパク質の相同性を評価できる。比較した蛋白質間でのスコアが高いほど、これらの蛋白質は相同性が高いと評価できる。ローカルアライメントの場合、ダイナミックプログラミングによって配列の組み合わせを探索する際にスコアに閾値を設け、部分配列の組み合わせの最適化を行うことにより、ローカルアライメントが行われる (例えば、Gotoh, O., Pattern matching of biological sequences with limited storage, Comput. Appl. Biosci. 3, 17-20, 1987)。このローカルアライメント法を用いることにより、二つの蛋白質間で共有しているオリゴペプチド以外の部分における相同の構造を検索することができる。

蛋白質間の相互作用は恐らく進化の過程で保存されてきているのであろう。後述する大腸菌のベロ毒素とBcl-2の例は、種を超えて一つの機能が蛋白質間相互作用というものに保存され、保存されてきた故に、構造上類似のアミノ酸配列が存在すると考えられる。さらに、後述するアミロイドプレカーサー蛋白質やフォンビルブランド因子 [Von Willebrand Factor (VWF)] プレカーサー蛋白質のプロセッシング過程でも、蛋白質分解という機能が蛋白質間相互作用に保存され、構造上でも類似のアミノ酸配列が存在していると考えることができる。

前記の思想に基づいて創出された蛋白質間相互作用予測方法によれば、

計算機上で予測された結果に基づき、分子生物学の枚举主義で得られた事実の積み重ねから得られている生命像とは異なった総合的作用相互関係を基とし、従来は単独としてしか知られていなかった機能のネットワークなども予想可能となり、新しい生命像の描出が可能となる。

- 5 さらには、一つの生物内だけではなく、二つの生物間、例えばヒトとその病原微生物との間での蛋白質の相互作用の予測が可能となると、これまで知られていなかった病原性の解明にもつながる。

- 具体的には、前記蛋白質間相互作用予測方法の態様の1つは、ゲノム
10 解析やcDNA解析により得られた機能未知の蛋白質について、又は機能が既知の蛋白質について、それが相互作用する相手の蛋白質を蛋白質データベースなどから抽出し予測する方法であって、例えば、以下のステップ(1)から(4)を構成要件としている。

- ステップ(1)において特定の蛋白質又はポリペプチドのアミノ酸配
15 列をある長さの複数個のオリゴペプチドに配列情報として分解し、ステップ(2)においてこのオリゴペプチドを共有する蛋白質又はポリペプチドを求め、ステップ(3)でそれら蛋白質間での部分構造の相同性をローカルアライメントにより評価し、ステップ(4)でさらに共有されているオリゴペプチドの評価を出現頻度から行う。

- 20 以下、各ステップを詳しく説明する。

ステップ(1):

- ゲノム解析やcDNA解析により得られた機能未知の蛋白質もしくは
ポリペプチド、又は機能が既知の蛋白質もしくはポリペプチドなどのある特定のタンパク質もしくはポリペプチド(A)について、そのアミノ
25 酸配列をアミノ末端からカルボキシル末端まで順に1アミノ酸残基ずつ
ずらしながらオリゴペプチドに分解し、配列情報とする。

例えば図1は、大腸菌O157:H7のベロ毒素2(VTII)のアミ

ノ末端から最初の20残基分〔図1の(a)〕についてアミノ酸長5個のオリゴペプチドに配列情報として分解したもの〔図1の(b)〕である。以下、アミノ酸及びオリゴペプチドはアミノ酸を一文字表記することにする。

- 5 ステップ(1)を実施する場合、配列情報として分解するオリゴペプチドのアミノ酸長は4~15個、好ましくは4~8個が適当である。オリゴペプチドの長さが長いほど、そのオリゴペプチドの特異性は、実施例2に示すように、上がる。

ステップ(2):

- 10 このステップではステップ(1)で分解して得られたオリゴペプチドのアミノ酸配列を持った蛋白質もしくはポリペプチド(C)を、又はこのオリゴペプチドと相同なアミノ酸配列を持った蛋白質もしくはポリペプチド(D)を、蛋白質もしくはポリペプチドのアミノ酸配列データベース中で検索する。検出された蛋白質もしくはポリペプチドC又はDは、
15 用いるオリゴペプチドにより数多く存在する場合もあるし、1個の場合もある。

例えば、図2は、大腸菌O157:H7のペロ毒素2(VTII)のアミノ酸末端から13残基分について分解された9種類のアミノ酸5個のオリゴペプチドについて、このオリゴペプチドを持つ蛋白質を蛋白質データベースSWISS-PROT version 35中で検索した
20 結果を示す。ペロ毒素2の場合、ヒトに食中毒や腎臓障害をおこすため、相互作用する相手の蛋白質として調べる対象となるのはヒトの蛋白質である。例えば、図2の2番目に示したオリゴペプチドKCILFを持つヒトの蛋白質は β アドレナリン受容体キナーゼ2[BETA-ADRE
25 NERGIC RECEPTOR KINASE 2 (ARK2 HUMAN)]であることが検索結果として得られる。

ステップ(3):

前記Aとステップ(2)の検索で得られたC又はDとの間でローカルアライメントを行う。

例えば、図3は、ペロ毒素2(VTII)と β アドレナリン受容体キナーゼ2(ARK2)とのローカルアライメントの結果を図示している。

5 ステップ(4):

ステップ(3)のローカルアライメントの結果から、前記Aと検出されたC及び/又はDとの間での部分配列の相同性があれば、C及び/又はDはAと相互作用する蛋白質又はポリペプチドBであると予測する。さらにAと検出されたC及び/もしくはDとが共有するオリゴペプチド
10 について、蛋白質データベースからアミノ酸の頻度及び/もしくはオリゴペプチドの頻度を計算し、そのオリゴペプチドの、蛋白質データベース中での、又はAもしくは検出されたC及び/もしくはDを持つ生物のゲノム中での、各特異性を評価し、特異性が高ければC及び/もしくはDはAと相互作用する蛋白質もしくはポリペプチドBであるとの予測は
15 より確度の高いものであると評価する。

例えば、前記オリゴペプチドK C I L Fは図4で示されている大腸菌ゲノムでコードされる全蛋白質中でのアミノ酸頻度〔図4の(a)〕から計算される組成比〔図4の(b)〕から、その特異性の指標は1284.86 $\times 10^{-10}$ と計算され、特異性が高い。ちなみに、大腸菌ゲノム中
20 で特異性が低いと計算されるアミノ酸配列長が5個のオリゴペプチドはL L L L Lであり、その特異性の指標は136344.34 $\times 10^{-10}$ である。また、一番特異性が高いアミノ酸配列長が5個のオリゴペプチドはC C C C Cであり、その特異性の指標は2.208 $\times 10^{-10}$ であるが、このオリゴペプチドは大腸菌ゲノム中には存在しない。したがって、ペロ毒素2と β アドレナリン受容体キナーゼ2とが相互作用すると
25 という予測は、特異性の指標の値から、確度が高いと評価される。

蛋白質間の相互作用をさらに確認するには、得られた相互作用する蛋白質の情報を基にその遺伝子をクローニングし発現させて、実験的に確認すればよい。例えば、実施例に後述するように、前記予測方法又は該予測方法のプログラムを担持した予測装置により相互作用すると予測された、V T IIとB c 1-2との相互作用は、B c 1-2を発現させた細胞と発現していない細胞を用い、これらをV T IIで処理後に、抗B c 1-2抗体及び抗V T II抗体を使用して共免疫沈降することで、実験的に確認することができた。さらに、例えば相互作用部位であると予測されるオリゴペプチドのアミノ酸配列に変異などを公知の方法で導入し、相互作用出来なくなることを確認することで、該オリゴペプチドが重要な相互作用部位であることを特定することもできる。実験的な相互作用の確認方法は、前記方法に限定されるものではなく、当業者の技術範囲で適用できるものであればいずれの手法を用いてもよい。

またさらに、相互作用部位であると予測されたオリゴペプチドが、蛋白質間相互作用を阻害し、該蛋白質の機能や作用を抑制することを確認すれば、このようなオリゴペプチドは、該タンパク質の作用の抑制剤として利用することができる。例えば、後述する実施例のように、V T IIとB c 1-2の相互作用部位であると予測されたオリゴペプチドN W G R Iは、V T IIによる細胞死を抑制し、抗細胞死剤として利用できる。このような低分子化合物は医薬品や試薬などとして利用が可能である。

次に、前記蛋白質間相互作用を予測する方法のプログラムを担持する記録媒体及び装置について説明する。前記記録媒体及び装置は、例えば、図5に構成例を示したように、少なくとも以下の(a)から(f)の手段を備えている。

入力手段(a):

ゲノム解析やcDNA解析により得られた機能未知の蛋白質もしくは

ポリペプチド、又は機能が既知の蛋白質もしくはポリペプチドなどのある特定のタンパク質もしくはポリペプチド（A）のアミノ酸配列情報を入力するための手段。

オリゴペプチド分解／記憶手段（b）：

- 5 入力手段（a）で入力されたアミノ酸配列情報をアミノ末端からカルボキシル末端へ順に1アミノ酸残基ずつずらしながら、ある長さの複数個のオリゴペプチドに配列情報として分解する手段と、その結果を記憶する手段。

記憶手段（c）：

- 10 蛋白質又はポリペプチドに関する入力されたデータベースを記憶する手段。

検索／記憶手段（d）：

- 15 記憶手段（c）に記憶された蛋白質もしくはポリペプチドに関するデータベースにアクセスし、前記オリゴペプチドのアミノ酸配列を持った蛋白質もしくはペプチド（C）を、又は前記オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を持った蛋白質もしくはポリペプチド（D）を、検索する手段とその結果を記憶する手段。

ローカルアライメント実行／記憶手段（e）：

- 20 前記Aと検出されたC及び／又はDとの間でローカルアライメントを実行する手段と、その結果を記憶する手段。

頻度計算／記憶表示手段（f）：

- 25 前記ローカルアライメントの結果、及び蛋白質又はポリペプチドのデータベースからアミノ酸の頻度及び／又はオリゴペプチドの頻度の計算を行いその結果から、蛋白質間相互作用の予測のための指標を計算する手段と、その結果を記憶し表示する手段。

さらに、前記蛋白質間相互作用予測プログラムにおいては、以下の（g）～（l）の手段を適宜に組み合わせて備えるようにすることもできる。

順位付け／記憶表示手段 (g):

- 検出された B が複数あるときにそれら B の間で、ローカルアライメントの結果、及び蛋白質データベースからアミノ酸の頻度及び／又はオリゴペプチドの頻度を計算して得られる結果を指標として、順位付けを行う機能と、その結果を記憶、表示する機能を有する手段。

位置表示手段 (h):

A と検出された B との間でアミノ酸部分配列間でアライメントされる場合、A と B のアミノ酸配列の全長を表示し、アライメントされる部分が全長上でどこにあるかを表示する手段。

- 10 立体構造計算／記憶表示手段 (i):

A もしくは検出された B の立体構造が既知の場合、又はホモロジーモデリングで立体構造モデルが作れる場合、その立体構造モデルを計算し、その立体構造上で、A と B との間でアライメントされるアミノ酸部分配列の構造を表示する手段。

- 15 蛋白質分類／記憶手段 (j):

検索範囲を絞るために蛋白質又はポリペプチドのデータベース中の蛋白質又はポリペプチドを物性、機能、及び／又は由来によって分類し、記憶する機能を持つ手段。

逐次入力手段 (k):

- 20 蛋白質又はポリペプチドのデータベース中の各蛋白質又はポリペプチドを A として逐次、入力する手段。

記憶手段 (l):

ゲノムデータベースを記憶する機能を持つ手段。

以上の手段は、適当な記録媒体に担持されて使用に供される。

25

また使用の一態様としては、これらの手段がプログラムとして選択的に装置内記録媒体に担持された装置として供給されてもよい。該蛋白質

間相互作用を予測する装置は、以下の様に作動する（図5を参照）。

入力手段（a）により入力された機能未知の蛋白質もしくはポリペプチド、又は機能が既知の蛋白質もしくはポリペプチドなどのある特定の蛋白質もしくはポリペプチド（A）は、オリゴペプチド分解／記憶手段
5 （b）により、そのアミノ酸配列情報がある長さの複数個のオリゴペプチドに分解され、該オリゴペプチドは記憶される。このとき、該Aの入力は、所望により、蛋白質又はポリペプチドに関して入力されたデータベースを記憶する手段（c）から、逐次入力手段（k）により、逐次入力される。前記記憶されたオリゴペプチドのアミノ酸配列について、手段
10 （c）に対して検索／記憶手段（d）により検索がなされ、該オリゴペプチドのアミノ酸配列を持った蛋白質もしくはポリペプチド（C）、又は該オリゴペプチドのアミノ酸配列と相同のアミノ酸配列を持った蛋白質もしくはポリペプチド（D）が検出されて記憶される。検索時、蛋白質分類／記憶手段（j）により、データベース中の蛋白質又はポリペ
15 チドを分類し検索範囲を絞って、その中から検索することもできる。前記Aと検出されたC又はDについてローカルアライメント実行／記憶手段（e）によりローカルアライメントがなされ、その結果が記憶される。次に、頻度計算／記憶表示手段（f）により、手段（c）に記憶されたデータベースからアミノ酸の頻度及び／又はオリゴペプチドの頻度の計
20 算がなされ、その結果と前記得られたローカルアライメントの結果から蛋白質間相互作用の予測のための指標が計算されて記憶される。そして、得られた、前記Aと相互作用すると予測される蛋白質もしくはポリペプチドC又はD、及びこれらの蛋白質間で共有されているオリゴペプチドのアミノ酸配列、該オリゴペプチドの頻度、並びに蛋白質間相互作用予
25 測のための指標などが装置の画面に表示される。表示された結果から、蛋白質間相互作用予測のための指標を基に、前記Aと相互作用を有する蛋白質もしくはポリペプチド（B）を得ることができる。また、前記B

について、手段（c）に記憶された蛋白質の機能情報や、所望により備えられたゲノムデータベースを記憶する手段（l）からの遺伝子情報など、が表示されることもできる。検出されたBが複数ある時には、順位付け／記憶表示手段（g）により、それらBの間で、前記Aと相互作用する特異性の高い順から順位付けがなされることができる。また、検出されたBと前記Aとの間でアライメントされるアミノ酸の部分配列が、該Aと該Bとの全長アミノ酸のどの部分であるかを位置表示手段（h）で表示することもできる。さらに、立体構造計算／記憶表示手段（i）により、該Aと該Bの立体構造を示し、AとBとの間でアライメントされる部分をも示すことができる。なお、これらの手段（a）～（l）以外にも、この装置は、図5にも例示したようなキーボード（m）、及び制御手段（n）、出力手段（o）などを備えたものであってもよい。

前記蛋白質間相互作用の予測方法又は前記予測装置はさらに、特定の蛋白質又はポリペプチド（A）と蛋白質又はポリペプチド（B）との蛋白質間相互作用を調整する新規化合物を選別する方法に使用することができる。前記蛋白質又はポリペプチド（A）と蛋白質又はポリペプチド（B）との蛋白質間相互作用を調整する新規化合物の選別方法は、キーとなるオリゴペプチドのアミノ酸配列の情報を基に行われる。選択されたオリゴペプチドのアミノ酸配列、又はその相同なオリゴペプチドのアミノ酸配列、又はそのアミノ酸配列もしくは相同なアミノ酸配列を含むポリペプチドは、そのものがAとBとの蛋白質間相互作用に調整機能を発揮する可能性がある。例えばAに対して、レセプター機能を有するBが存在する場合は、前記手法により選別されるオリゴペプチドはAとBの蛋白質間相互作用に対して拮抗性を有する可能性が高い。又は、例えばAがBとの相互作用で活性化されるような場合、前記手法により選別されるオリゴペプチドはアゴニストとしての機能を持つ可能性が高い。

具体的には、下記実験例に詳述するように、本発明により相互作用すると予測され、かつ実験的にその相互作用が確認された、V T IIとB c l-2とが共有する配列表の配列番号1に記載のオリゴペプチドNWGR Iは、V T IIとB c l-2との相互作用による複合体形成を阻害し、
5 さらにV T IIによる細胞死誘導を抑制することが実験的に確認された。このことから、NWGR Iオリゴペプチドは、V T IIによる細胞死誘導に係る疾病を調節する薬剤として、例えばV T IIを有する大腸菌O 1 5 7による疾病の治療薬、より具体的には抗細胞死剤として使用できる可能性がある。また、該アミノ酸配列と相同なアミノ酸配列を有するオリ
10 ゴペプチドであって、V T IIとB c l-2との相互作用を調整する機能を有するオリゴペプチド、又は該アミノ酸配列もしくは該アミノ酸配列と相同なアミノ酸配列を含むポリペプチドであって、V T IIとB c l-2との相互作用を調整する機能を有するポリペプチドも、V T IIによる細胞死誘導に係る疾病を調節する薬剤として使用できる。さらに、これ
15 らオリゴペプチド及びポリペプチドを利用してV T IIとB c l-2との相互作用を調整する機能を有する新規化合物を、ドラッグデザインすることにより、又は公知のスクリーニング法を適用することにより得ることもできる。

20 このように、本発明の1つの態様である前記選別方法によって得られるオリゴペプチドの情報を基にしてドラッグデザインされる新規化合物は、特定の蛋白質又はポリペプチド(A)と蛋白質又はポリペプチド(B)との蛋白質間相互作用の調整機能を有する。すなわち、自体公知のドラッグデザイン手法により、前記選別方法で得られるオリゴペプチドの誘
25 導体、及び該オリゴペプチドと相同な構造を持つ低分子化合物が、前記Aと前記Bとの蛋白質間相互作用を予測することによって構築される。

なお、前記予測方法は、特定の蛋白質又はポリペプチド(A)と蛋白

質又はポリペプチド（B）との蛋白質間相互作用に係るオリゴペプチドをコードするオリゴヌクレオチドの配列決定方法としても極めて有用であり、この情報を基に、置換、欠失、付加、挿入又は誘導変異などの自体公知の方法を応用して、有用なオリゴヌクレオチドを得ることができる。得られたオリゴヌクレオチドは、前記Aと前記Bとの蛋白質間相互作用を遺伝子レベルで調節するための化合物を得るために使用できる。例えば、蛋白質間相互作用を抑制するための、アンチセンスオリゴヌクレオチドなどの構築に利用することができる。また、得られたオリゴヌクレオチドは、蛋白質間相互作用が関係している疾病において、その診断に用いることもできる。

また、本発明の1つの態様は、蛋白質間相互作用の予測方法又は予測装置により求められた、蛋白質間相互作用があると予測されるヒト由来蛋白質の組み合わせ群である。該蛋白質の組み合わせ群はカタログ化及びデータベース化されて提供され得る。この蛋白質間相互作用を有する蛋白質の組み合わせ群から、疾病に関与する可能性がある既知蛋白質の情報を基に、疾病に関与する蛋白質間相互作用を有するものを選択することにより、疾病に関与する蛋白質間相互作用を有する蛋白質の組み合わせ群が得られる。これらはカタログ化及びデータベース化されて提供され得る。これらの蛋白質の組み合わせは、疾病を治療もしくは予防するための薬剤として、又は薬剤を得るための手段として有用である。例えば、得られた蛋白質の組み合わせを利用して、二つの蛋白質間の相互作用を調整できる化合物を、公知のスクリーニング方法を用いて探索して得ることができる。

蛋白質間相互作用を有する二つの蛋白質の組み合わせの中でも、特に一方の蛋白質が蛋白質プロセシング機能を有する酵素であり、他方の蛋白質を切断する場合、その切断される蛋白質のプロセシング部位の予測

を、前記蛋白質間相互作用の予測方法又は予測装置により行うことができる。

具体例として下記実施例に示したように、アミロイドプレカーサー
ロテインとそのプロセシングに關与する酵素との相互作用、及びフォン
5 ビルブランド因子プレカーサー蛋白質とそのプロセシングに係る酵素フ
ューリン (f u r i n) との相互作用を予測することができた。すなわ
ち、前記蛋白質間相互作用の予測方法又は予測装置は、プレカーサー蛋
白質が切断されて成熟蛋白質として機能するときの切断部位の予測が可
能である。このように、蛋白質間相互作用を予測することにより、従来
10 知られていなかった、疾病に關与する蛋白質プロセシング作用を有する
酵素並びに該酵素により切断される蛋白質を得ることができる。

実施例

本発明の利点、特徴、及び可能な適用については、例示的な実施形態
15 を参照しながら以下に更に詳細に記載するが、本発明は以下の例によ
って限定されるものではない。また、本実施例においては、蛋白質デー
タベースとして S W I S S - P R O T v e r s i o n 3 5 を用いたが、
他の蛋白質データベースなどを用いてもよい。

20 (実施例 1)

図 1 は、ステップ (1) の実施例として、大腸菌 O 1 5 7 : H 7 のベ
ロ毒素 2 (V T II) のアミノ末端から最初の 2 0 残基分〔図 1 の (a)〕
についてアミノ酸配列長が 5 個のオリゴペプチドに分解したもの〔図 1
の (b)〕である。また、図 1 の (c) は、ベロ毒素 2 (V T II) のアミ
25 ノ末端から最初の 2 0 残基分についてアミノ酸配列長が 6 個のオリゴペ
プチドに分解したものである。

(実施例 2)

前記蛋白質間相互作用予測方法のステップ (4) において、蛋白質又はポリペプチド間相互作用の予測の指標として、蛋白質又はポリペプチドのデータベースからアミノ酸の頻度及びオリゴペプチドの頻度から計算される数値を用いる。ここではその実施例として、図 4 で示されている大腸菌ゲノムでコードされる全蛋白質中でのアミノ酸頻度から計算されるオリゴペプチドの特異性を挙げる。20 種類あるアミノ酸の大腸菌ゲノムでコードされる全蛋白質中での出現頻度 [図 4 の (a)] から、各アミノ酸 a_i の組成比率 A_i は図 4 の (b) のようになる。

- いま、特異性を求めたいオリゴペプチドが $a_1 a_2 a_3 a_4 a_5$ であるとき、このオリゴペプチドの特異性は $A_1 \times A_2 \times A_3 \times A_4 \times A_5$ と計算される。例えば、オリゴペプチドが K C I L F のとき、 $4.406610 \times 1.170608 \times 6.004305 \times 10.639652 \times 3.898962 \times 10^{-10}$ と計算される。また、オリゴペプチド L L L L の特異性は $136344.34 \times 10^{-10}$ 、オリゴペプチド C C C C の特異性は 2.20×10^{-10} と計算される。

- この実施例における特異性は数字が小さいほど特異性が高い。アミノ酸配列長が 5 個のオリゴペプチドで最も特異性が高いのはオリゴペプチド C C C C C であり、このオリゴペプチドは大腸菌ゲノムでコードされる全蛋白質中には出現しない。逆に最も特異性が低いのはオリゴペプチド L L L L L である。

ステップ (1) で、蛋白質又はポリペプチド (A) をオリゴペプチドに分解するとき、オリゴペプチドの長さが長いほど、そのオリゴペプチドの特異性は高くなる。

(実施例 3)

前記蛋白質間相互作用予測方法のステップ (4) において、蛋白質又

はポリペプチド間相互作用の予測の指標としてローカルアライメントの結果を用いるが、ここではその実施例として、Gotohの方法 (Gotoh, O., Pattern matching of biological sequences with limited storage, Comput. Appl. Biosci. 3, 17-20, 1987) による部分配列のアライメントのスコア (score) を使った場合を挙げる。また、以下の本実施例ではスコアが25.0以上のとき、蛋白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との間でアミノ酸部分配列がアライメントされる (相同である) とした。

10 蛋白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との間で、 m 個のアミノ酸部分配列がアライメントされ、それらのスコアが S_i ($1 \leq i \leq m$)、Bのアミノ酸配列の長さを L_B とする。このローカルアライメントの結果から計算されるAとBとの間の相互作用の予測の指標を $\text{sum}(S_i) / L_B$ と定義する。この指標は高いほど、相互作用が強いと評価する。

15

(実施例4)

(VTIIとBcl-2との相互作用予測)

大腸菌O157:H7のベロ毒素2 (VTII) はヒトに食中毒や腎臓障害をおこすが、その作用機序は良く分かっていない (Sandvig, K., et al., Exp. Med. Biol. 412, 225-232, 1997; Paton, J.C., and Paton, A.W., Clin. Microbiol. Rev. 11, 450-479, 1998)。

20 また、この蛋白質は毒性を有する蛋白質であるから、細胞死に関係したヒトの蛋白質は、この蛋白質が相互作用する蛋白質の候補となる。そこで、この蛋白質が相互作用するであろうヒトの蛋白質を蛋白質データベースSWISS-PROT version 35中から、検索範囲を

25

細胞死に関係したヒトの蛋白質に絞って探し出し、それらが実際に相互作用していることを確認した例を以下に示す。

ベロ毒素2のアミノ酸配列長5個のオリゴペプチドで、細胞死に関係したヒトの蛋白質と共有されるのは、図6にその検索結果が示されているように、SWISS-PROT version 35においてLC
5 LLL、QRVAA、EFSGN、NWGRIの4つである。(図6中ではヒトの蛋白質はSWISS-PROT version 35の蛋白質IDで示されている。)これらのオリゴペプチドの特異性を、図4で示されている大腸菌ゲノムでコードされる全蛋白質中でのアミノ酸頻度から計算すると、LC
10 LLLの特異性は 15001.03×10^{-10} 、またQRVAA、EFSGN、NWGRIについては、それぞれ特異性は 15584.55×10^{-10} 、 3801.65×10^{-10} 、 1479.85×10^{-10} となる。したがって、これら4つのオリゴペプチドの中で最も特異性が高いのはNWGRIである。

15 オリゴペプチドNWGRIはベロ毒素2とBcl-2、Bcl-xL、MCL-1という3つのヒト蛋白質との間で共有されている。ベロ毒素2(VTH)とBcl-2、Bcl-xL、MCL-1でローカルアライメントを実行すると、その結果はそれぞれ図7、8、9のようになり、部分的にアミノ酸配列上で相同性を持つ。そこで、実施例3のように、
20 ローカルアライメントのスコアの合計値を各蛋白質の長さで割った指標を計算すると、

$$\text{Bcl-2} (30.0 + 27.0 + 25.0) / 239 = 0.343$$

$$\text{Bcl-xL} (30.0 + 29.0 + 27.0) / 233 = 0.369$$

$$\text{MCL-1} (34.0 + 30.0 + 28.0 + 26.0) / 350 = 0.337$$

25 となる。

これら3つの蛋白質のうち、Bcl-2とBcl-xLは同じファミリーを構成しているので、Bcl-2及びBcl-xLとMCL-1と

で、どちらがペロ毒素2と相互作用しやすいか、との判断は前記方法ではこのローカルアライメントから計算される指標を基に、B c l - 2 及び B c l - x L となる。

大腸菌 O 1 5 7 : H 7 のペロ毒素にはペロ毒素2の他にペロ毒素1 (V T I) がアイソフォームとして存在する。ペロ毒素1の毒性はペロ毒素2に比べて弱く、ペロ毒素2の50分の1ほどである (T e s h , V L . , e t a l . , 1 9 9 3 , I n f e c t . I m m u n . 6 1 , 3 3 9 2 - 3 4 0 2) 。蛋白質データベース S W I S S - P R O T v e r s i o n 3 5 において、ペロ毒素1とアミノ酸長5個のオリゴペプチドを共有する細胞死に関係したヒトの蛋白質は P 2 X 1 _ H U M A N であり、その共有されるオリゴペプチドは S S T L G である。しかし、このオリゴペプチド S S T L G の特異性は、図4で示された大腸菌ゲノムでコードされる全蛋白質中でのアミノ酸頻度から計算すると、 14385.63×10^{-10} であり、N W G R I と比べ特異性は約10倍低い。

また、ペロ毒素1において、ペロ毒素2のアミノ酸配列長5個のオリゴペプチド N W G R I に対応するオリゴペプチドは N W G R L であるが、この特異性は図4から 2622.30×10^{-10} であり、N W G R I と比べて低い。B c l - 2 及び B c l - x L もアミノ酸配列長4個のオリゴペプチド N W G R をペロ毒素1と共有するが、N W G R I と N W G R L の特異性の比較から、B c l - 2 及び B c l - x L はペロ毒素1よりもペロ毒素2に強く相互作用することが予測される。また、ペロ毒素1と B c l - 2 又は B c l - x L とのローカルアライメントの結果 (図10) から計算される指標もそれぞれ $(27.0 + 26.0) / 239 = 0.222$ 、 $26.0 / 233 = 0.112$ (N W G R 部分以外に相同なアミノ酸部分配列はない) となり、ペロ毒素1と B c l - 2 又は B c l - x L との相互作用は、ペロ毒素2と B c l - 2 又は B c l - x L との相互作用よりもかなり弱いと判断される。

(実施例 5)

(V T II と B c l - 2 との相互作用予測の実験的確認)

実施例 4 でペロ毒素 2 とヒトの B c l - 2 又は B c l - x L とが相互作用するという予測は確度が高いと評価された。この予測結果を基に、
5 ペロ毒素 2 と B c l - 2 が実際に相互作用することを実験的に確認した〔図 1 1 の (a) 及び図 1 1 の (b) 右図〕。実験的にはヒトの肝臓ガン細胞 (本来 B c l - 2 遺伝子を発現していない) H e p G 2 、及びこの細胞に B c l - 2 発現ベクターを導入し B c l - 2 を発現するようにした細胞 B 1 0 について、ペロ毒素 2 (V T II) を作用させ、抗 B c l -
10 2 抗体 (B c l - 2 I P s) 及び抗 V T II 抗体 (V T II I P s) を用いて自体公知の方法で共免疫沈降を行った。

図 1 1 の (a) 左図は共免疫沈降後に抗 B c l - 2 抗体を用いてウエスタンブロット分析した結果、図 1 1 の (a) 右図は抗 V T II 抗体を用いてウエスタンブロット分析した結果である。これらの図から B 1 0 細胞において V T II と B c l - 2 との複合体が共免疫沈降していること、
15 すなわちこれら 2 つの蛋白質が相互作用していることが確認できた。また、B 1 0 細胞にペロ毒素 1 (V T I) 又はペロ毒素 2 (V T II) を作用させ、細胞内のどのフラクションからこれらの蛋白質が検出されるかを、抗 V T I 抗体及び抗 V T II 抗体を用いて確認した。細胞死にはミト
20 コンドリアにある B c l - 2 が非常に重要な役割を果たしているが、ペロ毒素 2 (V T II) はミトコンドリアフラクションからも検出される〔図 1 1 の (b) 右図〕。

一方、ペロ毒素 1 はミトコンドリア B c l - 2 とは強い相互作用を示さないことが、ミトコンドリアフラクションからこの蛋白質が検出されないことから実験的に証明された。図 1 1 の (b) 左図にその結果を示す。
25

(実施例 6)

ペロ毒素 2 (VTII) と Bc1-2 のアミノ酸配列の全長を表示し、アライメントされる部分配列が全長上でどこにあるかを表示した実施例を図 12 に示す。

5

(実施例 7)

Bc1-xL の立体構造は既知であり、蛋白質立体構造データベース PDB にその構造が登録されている。この構造を基に Bc1-xL の立体構造上でペロ毒素 2 と部分アミノ酸配列が相同な部分を図 8 のローカルアライメントの結果に基づき図 13 中に太線で示す。

10

(実施例 8)

ペロ毒素 2 はこれまでリボゾーム RNA の一部を切断することにより蛋白質合成を停止させ、その毒性を発揮すると考えられてきた。リボゾーム RNA の一部を切断するリシンという蛋白質の立体構造が蛋白質立体構造データベース PDB に登録されており、その構造を基にペロ毒素 2 のホモロジーモデリングを行い、モデル立体構造上で、Bc1-xL と相同なアミノ酸部分配列の構造を図 8 のローカルアライメントの結果に基づき図 14 中に太線で示す。

15

20

(実施例 9)

(VTII による細胞死誘導の NWGR I による抑制)

次に実施例 4 で見出された、VTII と Bc1-2 とが共有しているオリゴペプチドである NWGR I が、VTII と Bc1-2 との相互作用を調整しうることを、実験的に確認した。まず、実施例 5 で使用した Bc1-2 を発現している B10 細胞の抽出物と、ビオチン化した VTII との複合体形成を、NWGR I オリゴペプチドの存在下で行い、ファウ

25

エスタンプロット法 (Far-Western blotting analysis) により解析した。NWGR Iオリゴペプチドは、濃度依存的にVT IIとBcl-2との複合体形成を阻害した。更に、B10細胞をNWGR Iオリゴペプチド0、10、50、100 μ Mで前処理し、10 ng/mlのVT IIとともに24時間処理し、アポトーシスによる細胞死誘導を測定した。合計約5000個の核をHoechst 33342/PI (Propidium iodide) で定法通り染色し、そのなかでアポトーシスを起こした核の比率を図15に示した。図に示すように、VT II単独の処理で約85%の細胞がアポトーシスによる細胞死を起こすのに対し、NWGR Iオリゴペプチドで前処理することにより、NWGR Iオリゴペプチドの濃度依存的に細胞死の誘導が抑制された。このことから、VT IIとBcl-2とが共有しているオリゴペプチドであるNWGR Iは、VT IIとBcl-2との相互作用を抑制してこれらによる複合体形成を阻害し、その結果VT IIによる細胞死誘導を抑制することが確認された。

(実施例10)

(CD4/gp120 HIV1)

ヒトエイズウイルスHIV1はヘルパーT細胞に感染する。このときウイルス表面の蛋白質gp120がヘルパーT細胞の表面蛋白質CD4に結合することがこの細胞への感染の第1段階として重要であることが分かっている。本実施例では、gp120とCD4の結合が前記予測方法で予測できるか確認した。蛋白質CD4のアミノ酸配列を長さ5個のオリゴペプチドに分解し、順次CD4由来のオリゴペプチドのアミノ酸配列を持つ蛋白質を蛋白質データベース中で検索すると、SLWDQというオリゴペプチドを共有する蛋白質としてgp120が抽出される〔図16の(a)〕。オリゴペプチドSLWDQは、SWISS-PRO

T version 35中でヒトの蛋白質としては蛋白質CD4にし
か存在せず、ヒトの蛋白質中では頻度が1であり、特異性が非常に高い。
また、このオリゴペプチド以外に、局所的に相同性の高い領域が存在す
る〔図16の(b)〕。CD4側からgp120と結合する際に、このオ
5 リゴペプチドSLWDQのすぐN末側の隣のアミノ酸残基アルギニン
(Arg)及び67-SFLTKGP-73が重要な役割を果たしてい
ることが分かっている(Kwong, P.D. et al., Nature,
vol. 398, 648-659, 1998)。また、gp120側から
CD4を認識するとき、図16の(b)で示した相同性の高い領域(2
10 89-KTIIVQLNETVKINCIRPNKKT-310)のす
ぐN末側の隣の数個のアミノ酸残基が、重要な役割を果たす領域の一つ
であることが分かっている(Kwong, P.D. et al., Natu
re, vol. 398, 648-659, 1998)。したがって、gp
120とCD4の結合が未知であったとしても前記予測方法で予測する
15 ことができる。

(実施例11)

(CED-4/MAC-1)

線虫(Caenorhabditis elegans)(C. ele
20 gans)は全ゲノム情報が明らかにされた最初の多細胞生物である。
このC. elegansについての実施例を一つここで示す。蛋白質C
ED-4はプログラム細胞死の制御に中心的な役割を果たす。MAC-
1はCED-4に結合し、細胞死を抑える蛋白質として発見された(W
u et al., Development, vol. 126, 9, 20
25 21-2031, 1999)。したがって、MAC-1とCED-4の結
合は既知であるが、本発明の検証のため、これら二つの蛋白質間でのオ
リゴペプチドの共有性を調べた。その結果、MAC-1とCED-4は

長さ5個のオリゴペプチドFPSVEを共有することが判明し、本発明が検証された。なお、C. elegansゲノム中のアミノ酸の頻度から計算されるこのオリゴペプチドの指数は5.436である。また、図17に図示するとおり、二つの蛋白質間では相同な部分領域が多く、これら5の蛋白質の結合性を強く示唆するものである（上側の配列がCED-4、下側の配列がMAC-1）。

（実施例12）

（APP/BASE）

10 APP (amyloid precursor protein: アミロイドプレカーサープロテイン) はアルツハイマー病の原因蛋白質の一つであり、この蛋白質が二ヶ所で切断されることにより、アルツハイマー病の原因となるアミロイドが生じる。この二ヶ所の切断箇所のうち、アミノ末端側を切断する酵素 (BASE) (bata secretase)
15 e) が最近発見された (VASSAR et al., Science, 286 (5440), 735-741, 1999)。BASEによるAPPの切断は、これら二つの蛋白質間で相互作用があることを意味する。そこで、本発明の検証のために、二つの蛋白質APPとBASEについてオリゴペプチドの共有性を調べた。APPとBASEは相同性の高い
20 アミノ酸5個のオリゴペプチドWYFDVとWYYEVを共有している。オリゴペプチドWYFDVは、SWISS-PROT version 35中でヒトの蛋白質としては蛋白質APPにしか存在せず、またWYYEVを持つヒトの蛋白質は未登録であり、どちらのオリゴペプチドも特異性が非常に高い。この結果により、本発明が検証された。図18に
25 二つの蛋白質間で相同な部分領域を図示する（上側の配列がAPP、下側の配列がBASE）。

(実施例 13)

(フューリンとフォンビルブランド因子)

フューリンは細胞内セリンプロテアーゼで、フォンビルブランド因子 (VWF)、アルブミン、補体 C3 などの分泌系の経路 (Pathway :
5 パスウェイ) に関係している。ここでは実施例としてフューリンと VWF との相互作用の例を挙げる。VWF はフューリンによりプレカーサー蛋白質から切断されて成熟蛋白質として機能する。フューリンによる VWF のプレカーサー蛋白質の切断には、これら二つの蛋白質間で相互作用が必要である。またフューリン自身もフューリンのプレカーサー蛋白質が切断されて成熟蛋白質となりプロテアーゼとして機能するようになる。
10 そこで、本発明の検証のために、フューリンプレカーサー蛋白質と VWF プレカーサー蛋白質のオリゴペプチドの共有性を調べた。二つの蛋白質はオリゴペプチド HCPPG を、フューリンプレカーサー蛋白質では 613-617 の位置に、VWF プレカーサー蛋白質では 1176-1180 の位置に持っている [図 19 の (a)]。どちらの位置も成熟蛋白質の領域内にある。オリゴペプチド HCPPG は SWISS-PROT version 35 でヒトの蛋白質としてはフューリンプレカーサー蛋白質及び VWF プレカーサー蛋白質にしか共有されておらず、頻度の観点から特異性が非常に高い。VWF プレカーサー蛋白質はフューリンによりアミノ酸残基の位置で 763 と 764 の間が切断される。
15 フューリンプレカーサー蛋白質と VWF プレカーサー蛋白質との間でローカルアライメントを行うと、フューリンによる VWF プレカーサー蛋白質の切断位置の近傍領域は、フューリンプレカーサー蛋白質に相同性の高い部分領域があることが分かる [図 19 の (b)]。すなわち、新規蛋白質が活性部分モチーフからプロテアーゼであると分かったとしても、相手の蛋白質、さらには相手の蛋白質中での切断位置が分からない場合、
25 本発明により、相手の蛋白質、さらには相手の蛋白質中での切断位置の

予測を行うことが可能である。

(実施例14)

(APPとPC7)

- 5 APPアルファは、アミロイドプレカーサープロテイン (amyloid precursor protein) (APP) がアミロイドを生じる二ヶ所の切断箇所とは異なる場所で切断されてできるペプチドである。APPアルファを生じるための切断にPC7 (proprotein convertase subtilisin/kexin type 7) が関与していることが最近発見された (Lopez-Perez E et al., J. Neurochem., vol. 73, 10 5, 2056-2062, 1999)。そこで、二つの蛋白質APPとPC7についてオリゴペプチドの共有性を調べると、APPとPC7は相同性の高い長さ6個のオリゴペプチドDSDPSGとDSDPNGを共有していた。オリゴペプチドDSDPSGは、SWISS-PROT version 35 15 中でヒトの蛋白質としては蛋白質APPにしか存在せず、またDSDPNGを持つヒトの蛋白質は未登録であり、どちらのオリゴペプチドも特異性が非常に高い。この結果により、本発明が検証された。図20に二つの蛋白質間で相同な部分領域を図示する。
- 20 図20のうちAPPの687-KLVFFAEDVGS-697のKとLの間がAPPアルファを生じるための切断箇所であり、これに相同な部分配列(359-RMPFYAEECAS-369)がPC7に存在することが図20に示されている。すなわち、本実施例は蛋白質の切断に関与する蛋白質を、本発明に従い予測することが可能であることを示している。
- 25

産業上の利用可能性

以上、詳しく説明したとおり、本発明によれば、ゲノム解析やcDNA解析により得られた機能未知の蛋白質、又は機能が既知の蛋白質が相互作用する相手の蛋白質について、蛋白質データベースを用いて、予測が可能となる。すなわち、近年充実してきたゲノム解析やcDNA解析に基づく蛋白質データベースを使用して、ゲノム情報の判明した一つの生物の中での、蛋白質間の相互作用の予測を計算機上行うことができる。計算機上での予測が可能になるため、two-hybrid法のような使用するcDNAライブラリーに結果が依存するハイリスクな手法を採らなくとも、計算機上での予測に基づき相互作用すると予測された蛋白質の情報を簡便に得ることができる。この予測が可能になったことにより、相互作用に関与するオリゴペプチドの配列が容易に推定でき、この情報を基に、新規な蛋白質間相互作用の調整機能をもった化合物のドラッグデザインも可能となった。本発明により、蛋白質間相互作用を解明するための実験が非常に効率化され、生化学、分子生物学、医薬品開発、農業、生物工学などの各種の分野で広く利用することができる。特に医薬品開発においては、今まで知られていなかった疾病メカニズムを予測することが可能となり、新しい医薬品が生まれる可能性を与える。

請求の範囲

1. 特定の蛋白質もしくはポリペプチド（A）と相互作用する蛋白質もしくはポリペプチド（B）を予測する方法であって、
 - （1）Aのアミノ酸配列をある長さの複数個のオリゴペプチドに配列情報として分解し、
 - （2）前記各オリゴペプチドのアミノ酸配列を持った蛋白質もしくはポリペプチド（C）を、又は前記オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を持った蛋白質もしくはポリペプチド（D）を、蛋白質もしくはポリペプチドのアミノ酸配列のデータベース中で検索し、
 - （3）前記Aと検出されたC又はDとの間でアミノ酸配列についてのローカルアライメントを行い、
 - （4）ローカルアライメントの結果、及び蛋白質もしくはポリペプチドのアミノ酸配列のデータベースにおけるアミノ酸の頻度及び／又はオリゴペプチドの頻度から計算される数値を指標として、検出されたC及び／又はDを、Aと相互作用する蛋白質もしくはポリペプチド（B）であると予測する、ことを特徴とする蛋白質間相互作用の予測方法。
2. 前記オリゴペプチドのアミノ酸配列の長さが、アミノ酸4～15個である請求の範囲第1項に記載の予測方法。
3. 特定の蛋白質もしくはポリペプチド（A）と相互作用する蛋白質もしくはポリペプチド（B）を予測する方法のプログラムを担持する記録媒体であって、少なくとも以下の（a）から（f）の手段を備えたことを特徴とする蛋白質間相互作用の予測プログラムを担持する記録媒体；
 - （a）Aのアミノ酸配列情報を入力し、記憶する手段と、
 - （b）この情報をある長さの複数個のオリゴペプチドに配列情報として分解する手段と、その結果の配列情報を記憶する手段と、
 - （c）入力された蛋白質データベースを記憶する手段と、
 - （d）該記憶された蛋白質データベースにアクセスし、前記オリゴペ

チドのアミノ酸配列を持った蛋白質もしくはポリペプチド (C) を、又は前記オリゴペプチドのアミノ酸配列と相同なアミノ酸配列を持った蛋白質もしくはポリペプチド (D) を、検出する手段と、その検出結果を記憶、計算する手段と、

- 5 (e) 前記 A と検出された C 又は D との間でローカルアライメントを行う手段と、その結果を記憶、計算する手段と、

- (f) 前記ローカルアライメントの結果値、及び蛋白質データベースからアミノ酸の頻度及び／又はオリゴペプチドの頻度の結果値を得、その結果値から、蛋白質間相互作用の予測のための指標を提示する手段と、
10 その結果を記憶し表示し、A と相互作用する B を検出する手段。

4. 請求の範囲第 3 項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；

- (g) 検出される B が複数ある場合、ローカルアライメントの結果値、及び蛋白質データベースにおけるアミノ酸の頻度及び／又はオリゴペ
15 プチドの頻度の結果値から計算される指標に基づき、該検出された複数の B の間での蛋白質間相互作用の強さの順位付けを行う手段と、その結果を記憶、表示する手段。

5. 請求の範囲第 3 項又は第 4 項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；

- 20 (h) ローカルアライメントにより、前記 A と前記検出された B との間でアミノ酸部分配列間でアライメントされる場合、当該 A と当該 B のアミノ酸配列の全長を表示し、アライメントされる部分配列が全長上でどこにあるかを表示する手段。

6. 請求の範囲第 3 項から第 5 項のいずれか 1 項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；
25

- (i) 前記 A もしくは前記検出された B の立体構造が既知の場合、又はホモロジーモデリングで立体構造モデルが作れる場合、その立体構造モ

デルを計算し、その立体構造上で、ローカルアライメントにより当該Aと当該Bの間でアライメントされるアミノ酸部分配列の構造を表示する手段。

7. 請求の範囲第3項から第6項のいずれか1項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；
- (j) 検索範囲を絞るために蛋白質データベース中の蛋白質を分類し、記憶する機能を備えた手段。
8. 請求の範囲第3項から第7項のいずれか1項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；
- (k) 蛋白質データベース中の各蛋白質を前記Aとして順次入力する手段。
9. 請求の範囲第3項から第8項のいずれか1項に記載の手段に加え、少なくとも以下の手段を備えたことを特徴とする記録媒体；
- (l) ゲノムデータベースを記憶する手段。
10. 請求の範囲第3項から第9項のいずれか1項に記載の記録媒体が担持する手段を備えた蛋白質間相互作用の予測装置。
11. 請求の範囲第1項又は第2項に記載の予測方法により特定の蛋白質又はポリペプチド(A)と相互作用すると予測された蛋白質又はポリペプチド(B)を得、ついで実験的にAとBとの間に相互作用があることを確認する、相互作用を有する蛋白質又はポリペプチドの特定方法。
12. 請求の範囲第10項に記載の予測装置により特定の蛋白質又はポリペプチド(A)と相互作用すると予測された蛋白質又はポリペプチド(B)を得、ついで実験的にAとBとの間に相互作用があることを確認する、相互作用を有する蛋白質又はポリペプチドの特定方法。
13. 請求の範囲第11項又は第12項に記載の、相互作用を有する蛋白質又はポリペプチドの特定方法で特定された蛋白質又はポリペプチド。
14. 請求の範囲第1項又は第2項に記載の予測方法を利用した、特定の蛋

白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との蛋白質間相互作用を調整する化合物の選別方法。

15 15. 請求の範囲第10項に記載の予測装置を用いた、特定の蛋白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との蛋白質間相互作用を調整する化合物の選別方法。

16. 請求の範囲第14項又は第15項に記載の選別方法によって得られる新規化合物。

10 17. 請求の範囲第14項又は第15項に記載の選別方法によって得られる化合物の情報を基にドラッグデザインして得られる、特定の蛋白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との蛋白質間相互作用を調整する機能を有する新規化合物。

18. ベロ毒素2 (VTII) とBcl-2との相互作用を調整する機能を有する配列表の配列番号1に示すアミノ酸配列を有するオリゴペプチド。

15 19. 配列表の配列番号1に示すアミノ酸配列を有するオリゴペプチドを含有する抗細胞死剤。

20. 請求の範囲第18項に記載のオリゴペプチドのアミノ酸配列と相同なアミノ酸配列を有するオリゴペプチドであって、VTIIとBcl-2との相互作用を調整する機能を有するオリゴペプチド。

20 21. 請求の範囲第18項又は第20項に記載のオリゴペプチドのアミノ酸配列を含むポリペプチドであって、VTIIとBcl-2との相互作用を調整する機能を有するポリペプチド。

22. 請求の範囲第18項もしくは第20項に記載のオリゴペプチド及び／又は請求の範囲第21項に記載のポリペプチドを利用する、VTIIとBcl-2との相互作用を調整する機能を有する化合物の選別方法。

25 23. 請求の範囲第1項もしくは第2項に記載の予測方法又は請求の範囲第10項に記載の予測装置を用いた、特定の蛋白質又はポリペプチド (A) と蛋白質又はポリペプチド (B) との蛋白質間相互作用に係るオリゴペ

プチドをコードするオリゴヌクレオチドの配列の決定方法。

24. 請求の範囲第1項もしくは第2項に記載の予測方法又は請求の範囲第10項に記載の予測装置により求められた、蛋白質間相互作用があると予測されるヒト由来蛋白質の組み合わせ群。
- 5 25. 請求の範囲第24項に記載の組み合わせ群から、疾病に關与する可能性がある既知蛋白質の情報を基にして選択を行う、疾病に關与する蛋白質間相互作用を有する蛋白質の組み合わせの選択方法。
26. 請求の範囲第25項に記載の方法で選択された、疾病に關与する蛋白質間相互作用を有する蛋白質の組み合わせ群。
- 10 27. 請求の範囲第26項に記載の組み合わせ群から選ばれるある組み合わせ及び／又は2蛋白質の相互作用を調整する化合物を選別する方法。
28. 請求の範囲第27項に記載の方法で得られた化合物。
29. 請求の範囲第1項もしくは第2項に記載の予測方法又は請求の範囲第10項に記載の予測装置を用いて、特定の蛋白質とこの蛋白質を切断する酵素との蛋白質間相互作用予測により、蛋白質のプロセシング部位を
- 15 予測する方法。
30. 請求の範囲第29項に記載の方法で得られた蛋白質プロセシング部位及び／又は該プロセシング部位と相同な部分配列を含むと予測されるアミノ酸配列を有するポリペプチド。

1/17

☒ 1

(a)

MKCILFKWVLCLLLGFSSVS

(b)

(c)

MKCIL

KCILF

CILFK

ILFKW

LFKWV

FKWVL

KWVLC

WVLCL

VLCLL

LCLLL

CLLLG

LLLGF

LLGFSS

LGFSS

GFSSV

FSSVS

MKCILF

KCILFK

CILFKW

ILFKWV

LFKWVL

FKWVLC

KWVLCL

WVLCLL

VLCLLL

LCLLLG

CLLLGF

LLLGFS

LLGFSS

LGFSSV

GFSSVS

2/17

図 2

MKCIL

なし

KCILF

BETA-ADRENERGIC RECEPTOR KINASE 2 (ARK2_HUMAN)

CILFK

WHITE PROTEIN HOMOLOG (WHIT_HUMAN)

ILFKW

なし

LFWV

なし

FKWVL

なし

KWVLC

なし

WVLCL

なし

VLCLL

GROWTH ARREST AND DNA-DAMAGE-INDUCIBLE PROTEIN GADD45
(GA45_HUMAN)

3/17

図 3

>Score = 32.0

ARK2 618 KCILFR 623

VTII 2 KCILFK 7

>Score = 27.0

ARK2 651 FKEAQRLLRRA 661

**_ **_*=*

VTII 193 FRQIQREFRQA 203

>Score = 26.0

ARK2 299 TEIILGLEHV 308

*** ****

VTII 44 TEISTPLEHI 53

>Score = 26.0

ARK2 379 CMLFK 383

VTII 3 CILFK 7

>Score = 25.0

ARK2 12 SYLMAMEKSKATPAARASKRI 32

*** ** * * = *** *

VTII 160 SYLALMEFSGNTMTRDASRAV 180

4/17

☒ 4

(a)

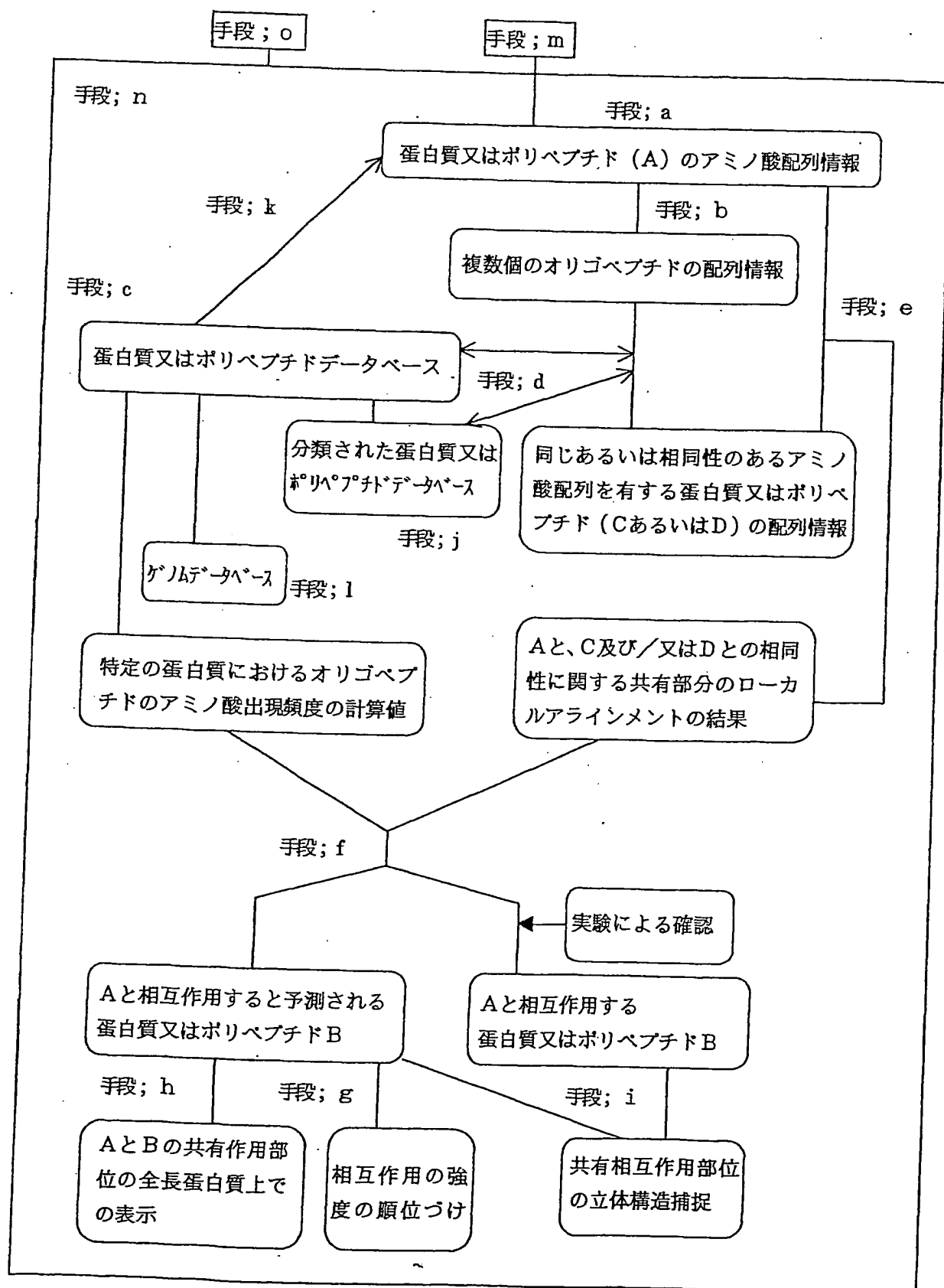
L: 144615
A: 128924
G: 100162
V: 95941
I: 81611
S: 79179
E: 78102
R: 75293
T: 73489
D: 69829
Q: 60214
P: 60182
K: 59895
N: 53727
F: 52995
M: 38745
Y: 38721
H: 30912
W: 20761
C: 15911

(b)

L: 10.639652%
A: 9.485230%
G: 7.369144%
V: 7.058596%
I: 6.004305%
S: 5.825378%
E: 5.746140%
R: 5.539476%
T: 5.406752%
D: 5.137477%
Q: 4.430080%
P: 4.427726%
K: 4.406610%
N: 3.952817%
F: 3.898962%
M: 2.850557%
Y: 2.848791%
H: 2.274266%
W: 1.527434%
C: 1.170608%

5/17

図 5



6/17

☒ 6

```

LCLLL    FASA_HUMAN
QRVAA    TRAI_HUMAN
EFSGN    DAPK_HUMAN
NWGRI    BCL2(BC2A)_HUMAN, BCLX_HUMAN, MCL1_HUMAN

```

☒ 7

>Score = 30.0

```

Bcl-2  143 NWGRI 147
      *****

```

```

VTII.  223 NWGRI 227

```

>Score = 27.0

```

Bcl-2  129 RFATVVEELFR 139
      ** ** * =*

```

```

VTII   182 RFVTVTAEALR 192

```

>Score = 25.0

```

Bcl-2  156 VMCVESVNREMSPLVDNIA 174
      * ** *** *_*=

```

```

VTII   36 VSSLNSIRTEISTPLEHIS  54

```

7/17

☒ 8

>Score = 30.0

Bcl-xL 136 NWGRI 140

VTII 223 NWGRI 227

>Score = 29.0

Bcl-xL 101 YRRAFSDLTSQLHITPG 117

=*== **

VTII 200 FRQALSETAPVYTMTPG 216

>Score = 27.0

Bcl-xL 5 NRELWVDF 12

==**= ***

VTII 22 SREFTIDF 29

8/17

☒ 9

>Score = 34.0

Mc1-1 326 GGIRNVLLAFAGVAGVGAG 344
* * *** * * ** *
VTII 225 GRISNVLPYRGEDGVRVG 243

>Score = 30.0

Mc1-1 260 NWGRI 264

VTII 223 NWGRI 227

>Score = 28.0

Mc1-1 275 AKHLKTINQES 285
** *_**=**
VTII 268 ARSVRAVNEES 278

>Score = 26.0

Mc1-1 134 LGKRPAVLPLLELVGESGNNTSTDGS 159
= *_== * = * *_= *** = * *
VTII 152 ISRHSLVSSYLALMEFSGNTMTRDAS 177

9/17

☒ 1 0

(a)

>Score = 27.0

Bcl-2 129 RFATVVEELFR 139

** ** * =*

VTI 182 RFVTVTAEALR 192

>Score = 26.0

Bcl-2 143 NWGRI 147

****=

VTI 224 NWGRL 228

(b)

>Score = 26.0

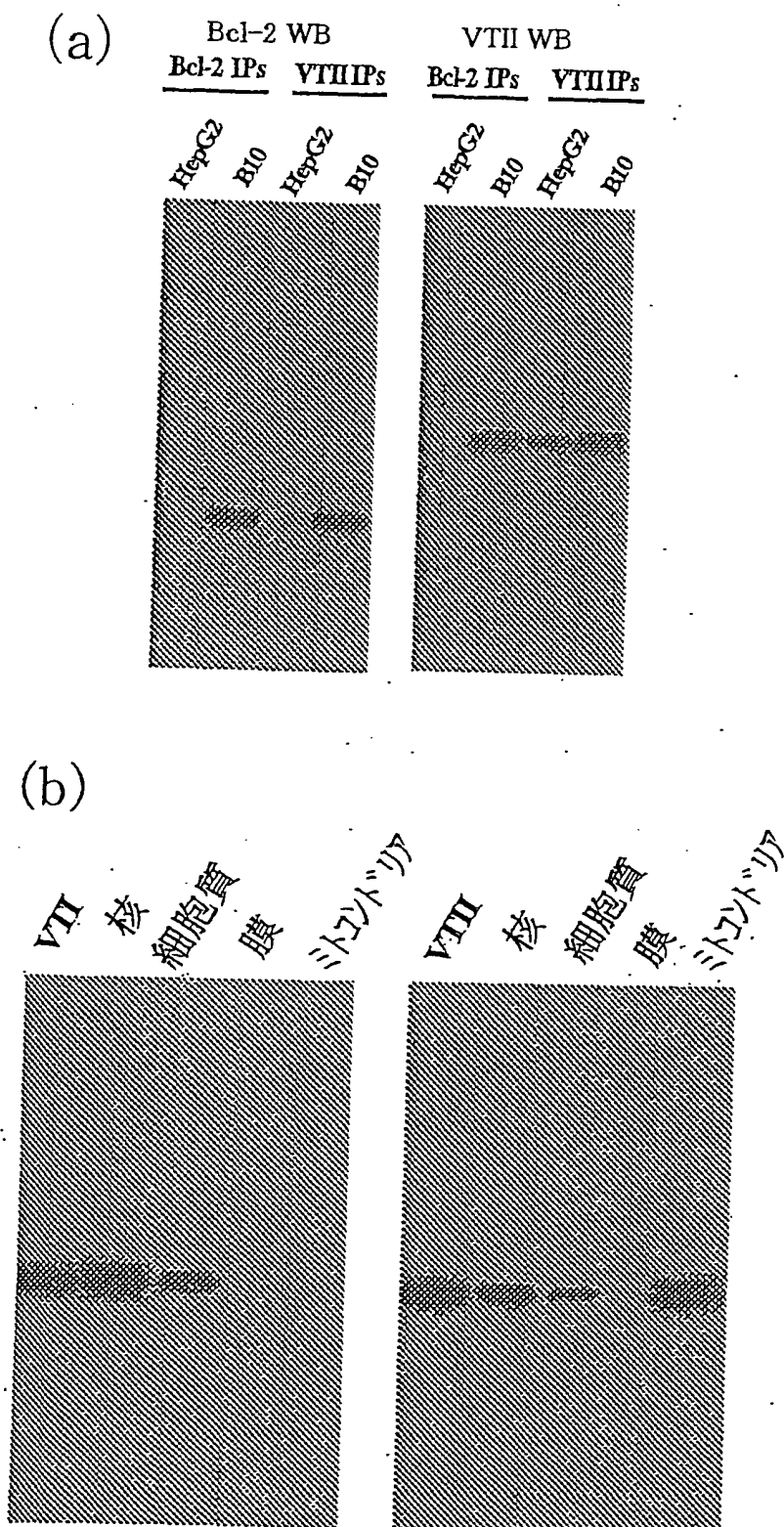
Bcl-xL 136 NWGRI 140

****=

VTI 224 NWGRL 228

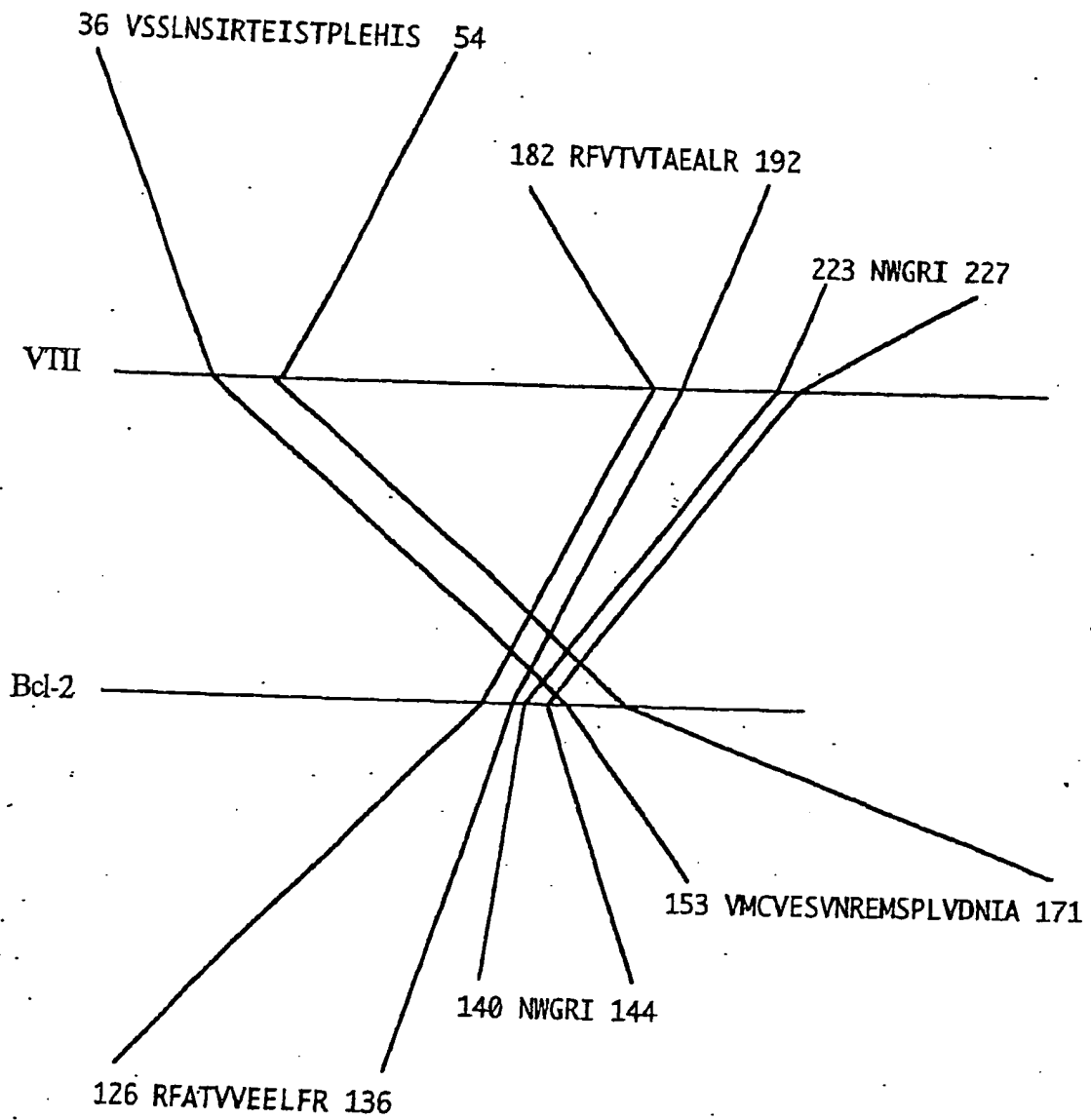
10/17

図 1 1



11/17

12



12/17

図 13

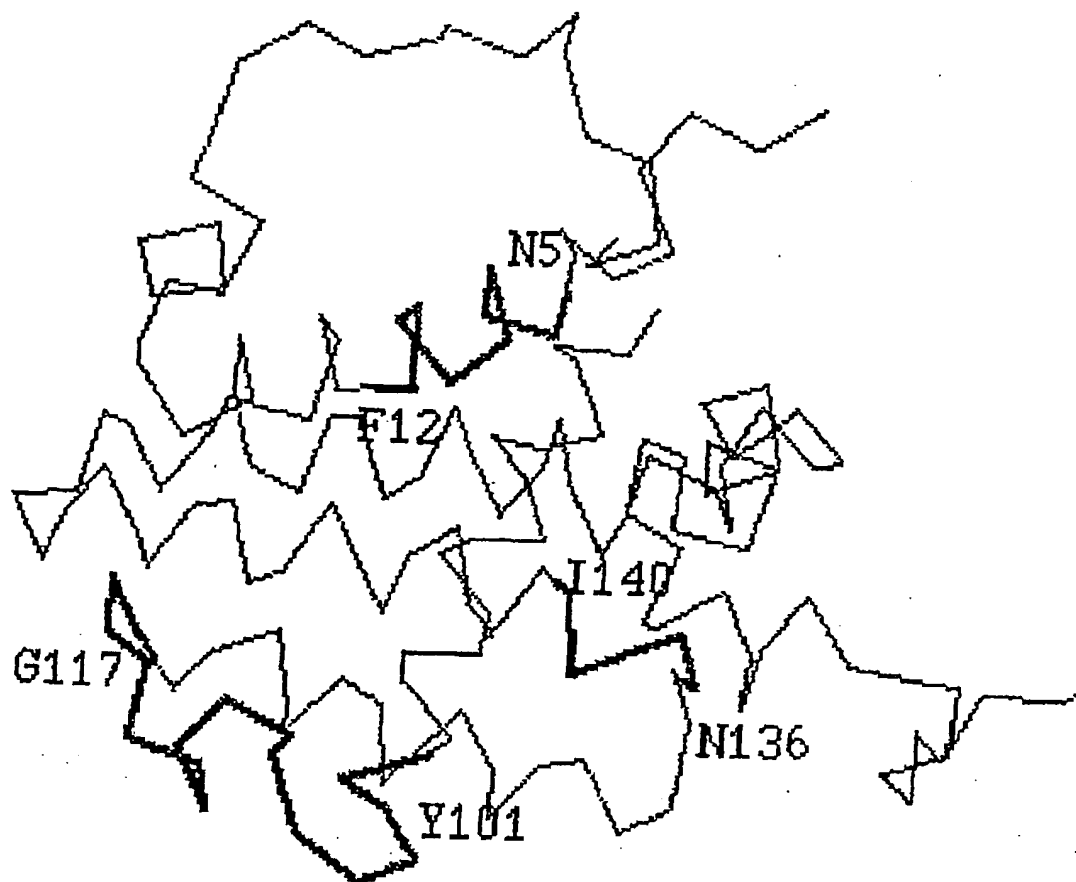
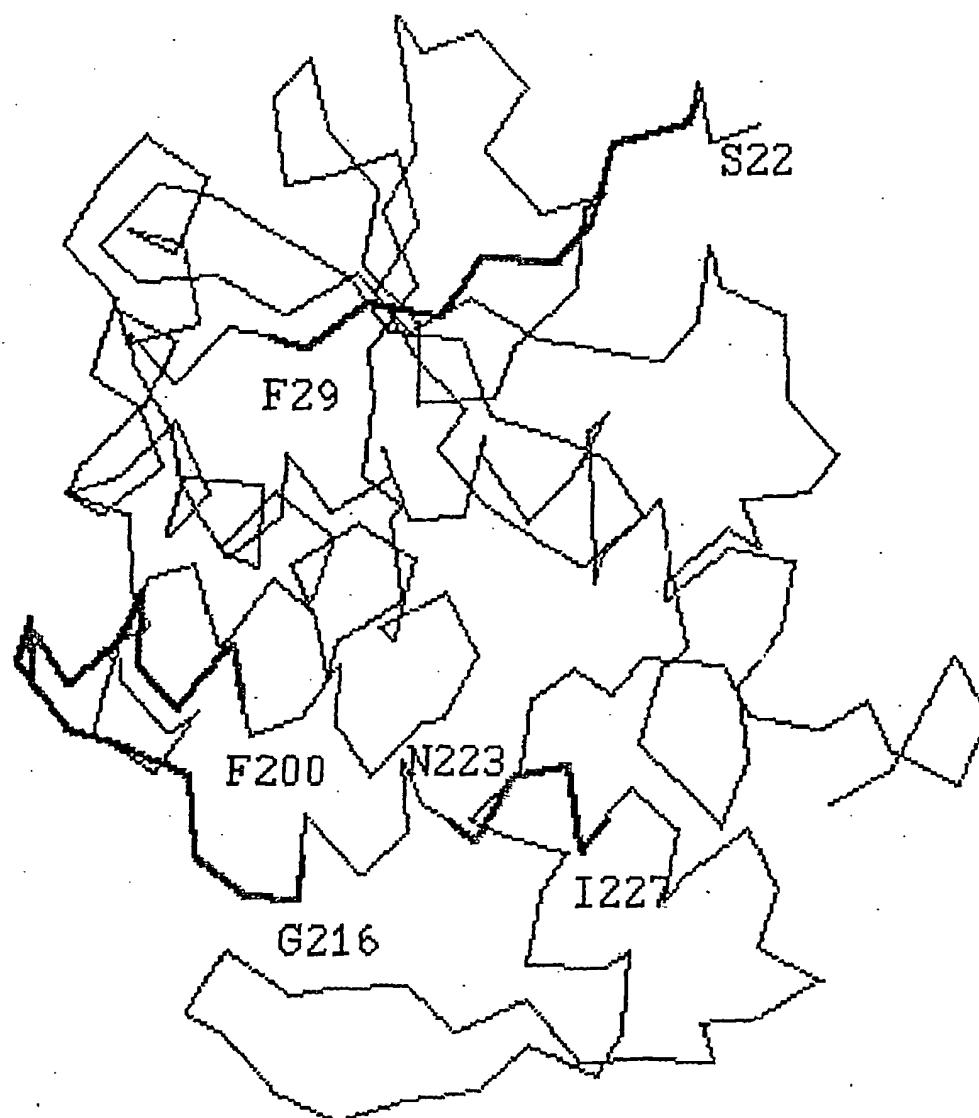


図 14



14/17

図 15

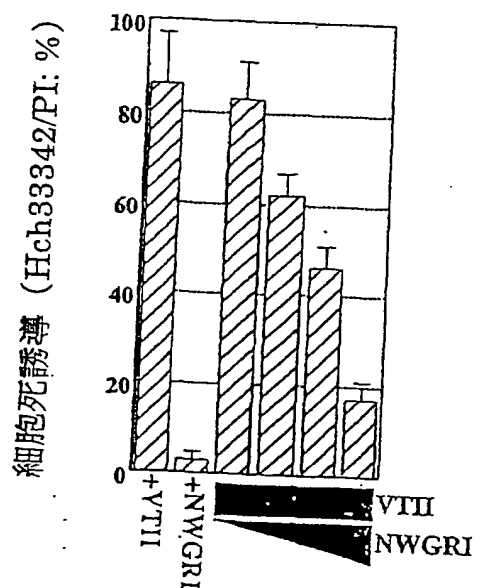


図 16

(a)

>Score = 29.0

CD4 85 SLWDQ 89

gp120 115 SLWDQ 119

(b)

>Score = 26.0

CD4 27 KVVLGKKGDTVELTCTASQKKS 48
* * = = *** = * = * =

GP120 289 KTIIVQLNETVKINCIRPNNKT 310

15/17

☒ 1 7

>Score = 30.0
 CED-4 158 HGRAGSGKSVIA 169
 ** * **=
 MAC-1 245 HGPPGCGKTMFA 256

>Score = 29.0
 CED-4 155 LFLHGRAGSGKSVIA 169
 ==* * * **=
 MAC-1 571 ILLCGPPGCGKTLA 585

>Score = 29.0
 CED-4 217 LLNFPSVE 224
 = *****
 MAC-1 698 FVDFPSVE 705

>Score = 27.0
 CED-4 93 FAINPDLL 101
 _ ****
 MAC-1 597 FSVKGPELL 605

>Score = 26.0
 CED-4 68 LGPLIDFFNYNNQSHLADFLEDYIDFAINEPDLL 101
 ** *** = ** * = *_* **
 MAC-1 723 LGEDIDFHEIAQLPELAGFTGADLAVFIHELSSL 756

>Score = 26.0
 CED-4 486 EIGNNNVSVPERHIPSHFQKFRRS 509
 _ * =** * = *** * *
 MAC-1 603 ELLNMYVGESERAVRTVFQRARDS 626

>Score = 25.0
 CED-4 301 FLEAYGMPM 309
 *** = * *
 MAC-1 215 FLEVCLAM 223

>Score = 25.0
 CED-4 428 DEVADRLKRLSK 439
 * * **** *
 MAC-1 381 DAVDGRLLRRTGR 392

16/17

☒ 1 8

```

>Score = 34.0
  APP 253 EEEEEAEPP 261
          * *** ***
  BASE 46 ETDEEPEEP 54

```

```

>Score = 27.0
  APP 307 WYFDV 311
          *****
  BASE 258 WYYEV 262

```

```

>Score = 26.0
  APP 50 GKWDSD 55
          *****
  BASE 135 GKWEGE 140

```

☒ 1 9

(a)

```

>Score = 34.0

furin-pre 613 HCPPG 617
          *****
VWF-pre 1176 HCPPG 1180

```

(b)

```

>Score = 25.0

furin-pre 73 TKRSLSPHRP 82
          =***** = *
VWF-pre 761 SKRSLSCRPP 770

```

17/17

図 2 0

```

>Score = 30.0
  APP 449 ERQQLVETHMA 459
      ***=* *= *=
  PC7 624 DRQRLLESAMS 634

```

```

>Score = 27.0
  APP 250 DGDEVEEEAEPP 261
      * ***** *= *
  PC7 730 DPDEVETESRGP 741

```

```

>Score = 26.0
  APP 44 HMNVQNGK 51
      ***** **
  PC7 773 HLDVPHGK 780

```

```

>Score = 26.0
  APP 53 DSDPSG 58
      *****=*
  PC7 563 DSDPNG 568

```

```

>Score = 26.0
  APP 755 NGYENPTY 762
      *** *= *
  PC7 340 DGYANSIY 347

```

```

>Score = 26.0
  APP 440 ESLEQEA 446
      *====*
  PC7 68 ETLEQQA 74

```

```

>Score = 26.0
  APP 687 KLVFFAEDVGS 697
      ** ***** *
  PC7 359 RMPFYAEECAS 369

```

```

>Score = 25.0
  APP 246 EDDGDEVEEEAE 259
      * * * * *==**
  PC7 62 EGDGEEETLEQQAD 75

```


1/1

SEQUENCE LISTING

<110> FUJITSU LIMITED

Daiichi Pharmaceutical Co., Ltd

<120> Method for estimating interaction between proteins and
recording medium carrying estimation programs therefor

<130> GP01-1001

<150> JP P2000-72485

<151> 2000-03-10

<160> 1

<170> PatentIn Ver. 2.1

<210> 1

<211> 5

<212> PRT

<213> Homo sapiens

<400> 1

Asn Trp Gly Arg Ile

1

5